Mon. Not. R. Astron. Soc. **350,** 1153–1173 (2004) doi:10.1111/j.1365-2966.2004.07744.x

Populating dark matter haloes with galaxies: comparing the 2dFGRS with mock galaxy redshift surveys

Xiaohu Yang,[1,2⋆] H. J. Mo,[2,3⋆] Y. P. Jing,[4] Frank C. van den Bosch[3] and YaoQuan Chu[1]

[1]*Centre for Astrophysics, University of Science and Technology of China, Hefei, Anhui 230026, China*
[2]*Department of Astronomy, University of Massachusetts, Amherst MA 01003-9305, USA*
[3]*Max-Planck-Institut für Astrophysik Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany*
[4]*Shanghai Astronomical Observatory; the Partner Group of MPA, Nandan Road 80, Shanghai 200030, China*

Accepted 2004 February 18. Received 2004 February 9; in original form 2003 December 15

ABSTRACT

In two recent papers, we developed a powerful technique to link the distribution of galaxies to that of dark matter haloes by considering halo occupation numbers as a function of galaxy luminosity and type. In this paper we use these distribution functions to populate dark matter haloes in high-resolution N-body simulations of the standard ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $\sigma_8 = 0.9$. Stacking simulation boxes of 100 h^{-1} Mpc and 300 h^{-1} Mpc with 512^3 particles each we construct mock galaxy redshift surveys out to a redshift of $z = 0.2$ with a numerical resolution that guarantees completeness down to $0.01L^*$. We use these mock surveys to investigate various clustering statistics. The predicted two-dimensional correlation function $\xi(r_p, \pi)$ reveals clear signatures of redshift space distortions. The projected correlation functions for galaxies with different luminosities and types, derived from $\xi(r_p, \pi)$, match the observations well on scales larger than $\sim$3 h^{-1} Mpc. On smaller scales, however, the model overpredicts the clustering power by about a factor two. Modelling the 'finger-of-God' effect on small scales reveals that the standard ΛCDM model predicts pairwise velocity dispersions (PVD) that are $\sim$400 km s^{-1} too high at projected pair separations of $\sim$1 h^{-1} Mpc. A strong velocity bias in massive haloes, with $b_{\rm vel} \equiv \sigma_{\rm gal}/\sigma_{\rm dm} \sim 0.6$ (where $\sigma_{\rm gal}$ and $\sigma_{\rm dm}$ are the velocity dispersions of galaxies and dark matter particles, respectively) can reduce the predicted PVD to the observed level, but does not help to resolve the overprediction of clustering power on small scales. Consistent results can be obtained within the standard ΛCDM model only when the average mass-to-light ratio of clusters is of the order of 1000 $(\rm M/L)_\odot$ in the B-band. Alternatively, as we show by a simple approximation, a ΛCDM model with $\sigma_8 \simeq 0.75$ may also reproduce the observational results. We discuss our results in light of the recent *WMAP* results and the constraints on σ_8 obtained independently from other observations.

Key words: methods: statistical – galaxies: haloes – dark matter – large-scale structure of Universe.

1 INTRODUCTION

The distribution of galaxies contains important information about the large-scale structure of the matter distribution. On large, linear scales the galaxy power spectrum is believed to be proportional to the matter power spectrum, therewith providing useful information regarding the initial conditions of structure formation, i.e. regarding the power spectrum of primordial density fluctuations. On smaller, non-linear scales the distribution and motion of galaxies is governed by the local gravitational potential, which is cosmology-dependent. One of the main goals of large galaxy redshift surveys is therefore to map the distribution of galaxies as accurately as possible, over as large a volume as possible. The Sloan Digital Sky Survey (SDSS; York et al. 2000) and the 2 degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) are two of the prime examples. These surveys, which are currently being completed, will greatly enhance and improve our knowledge of large-scale structure and will become the standard data sets against which to test our cosmological and galaxy-formation models for the decade to come.

However, two effects complicate a straightforward interpretation of the data. First of all, the distribution of galaxies is likely to be

⋆E-mail: xhyang@ustc.edu.cn (XY); hjmo@nova.astro.umass.edu (HJM)

biased with respect to the underlying mass density distribution. This bias is an imprint of various complicated physical processes related to galaxy formation such as gas cooling, star formation, merging, tidal stripping and heating, and a variety of feedback processes. In fact, it is expected that the bias depends on scale, redshift, galaxy type, galaxy luminosity, etc. (Kauffmann, Nusser & Steinmetz 1997; Jing, Mo & Börner 1998; Somerville et al. 2001; van den Bosch, Yang & Mo 2003a). Therefore, in order to translate the observed clustering of galaxies into a measure for the clustering of (dark) matter, one needs to either understand galaxy formation in detail, or use an alternative method to describe the relationship between galaxies and dark matter (haloes). One of the main goals of this paper is to advocate one such method and to show its potential strength for advancing our understanding of large-scale structure.

Secondly, because of the peculiar velocities of galaxies, the clustering of galaxies observed in redshift space is distorted with respect to the real-space clustering (e.g. Davis & Peebles 1983; Kaiser 1987; Regos & Geller 1991; Hamilton 1992; van de Weygaert & van Kampen 1993). On small scales, the virialized motion of galaxies within dark matter haloes smears out structure along the line-of-sight (i.e. the so-called 'finger-of-God' effect). On large scales, coherent flows induced by the gravitational action of large-scale structure enhance structure along the line-of-sight. Both effects cause an anisotropy in the two-dimensional, two-point correlation function $\xi(r_p, \pi)$, with r_p and π the pair separations perpendicular and parallel to the line-of-sight, respectively. The large-scale flows compress the contours of $\xi(r_p, \pi)$ in the π-direction by an amount that depends on $\beta \equiv \Omega_{\rm m}^{0.6}/b$. The small-scale peculiar motions imply that $\xi(r_p, \pi)$ is convolved in the π-direction by the distribution of pairwise velocities, $f(v_{12})$. Thus, the detailed structure of $\xi(r_p, \pi)$ contains information regarding the universal matter density $\Omega_{\rm m}$, the (linear) bias of galaxies b, and the pairwise velocity distribution $f(v_{12})$.

From the above discussion it is obvious that understanding galaxy bias is an integral part of understanding large-scale structure. One way to address galaxy bias without a detailed theory of how galaxies form is to model halo occupation statistics. One simply specifies halo occupation numbers, $\langle N(M)\rangle$, which describe how many galaxies on average occupy a halo of mass M. Many recent investigations have used such halo occupation models to study various aspects of galaxy clustering (Jing, Mo & Börner 1998; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; White 2001; Berlind & Weinberg 2002; Bullock, Wechsler, & Somerville 2002; Jing, Börner & Suto 2002; Kang et al. 2002; Marinoni & Hudson 2002; Scranton 2002; Zheng et al. 2002; Kochanek et al. 2003). In two recent papers, Yang, Mo & van den Bosch (2003, hereafter Paper I) and van den Bosch et al. (2003a, hereafter Paper II) have taken this halo occupation approach one step further by considering the occupation as a function of galaxy luminosity and type. They introduced the conditional luminosity function (hereafter CLF) $\Phi(L \mid M)\,{\rm d}L$, which gives the number of galaxies with luminosities in the range $L \pm {\rm d}L/2$ that reside in haloes of mass M. The advantage of this CLF over the halo occupation function $\langle N(M)\rangle$ is that it allows one to address the clustering properties of galaxies {as a function of luminosity}. In addition, the CLF yields a direct link between the halo mass function and the galaxy luminosity function, and allows a straightforward computation of the average luminosity of galaxies residing in a halo of given mass. Therefore, $\Phi(L \mid M)$ is not only constrained by the clustering properties of galaxies, as is the case with $\langle N(M)\rangle$, but also by the observed luminosity functions (LFs) and the halo mass-to-light ratios.

In Papers I and II we used the observed LFs and the luminosity- and type-dependence of the galaxy two-point correlation function to constrain the CLF in the standard ΛCDM cosmology. In this paper, we use this CLF to populate dark matter haloes in high-resolution N-body simulations. The 'virtual universes' thus obtained are used to construct mock galaxy redshift surveys with volumes and apparent magnitude limits similar to those in the 2dFGRS. This is the first time that realistic mock surveys have been constructed that (i) associate galaxies with dark matter haloes, (ii) are independent of a model of how galaxies form, and (iii) automatically have the correct galaxy abundances and correlation lengths as a function of galaxy luminosity and type. In the past, mock galaxy redshift surveys were constructed either by associating galaxies with dark matter particles (rather than haloes) using a completely {ad hoc} bias scheme (Cole et al. 1998), or by linking semi-analytical models for galaxy formation (with all their associated uncertainties) to the merger histories of dark matter haloes derived from numerical simulations (Kauffmann et al. 1999; Mathis et al. 2002).

We use our mock galaxy redshift survey to investigate a number of statistical measures of the large-scale distribution of galaxies. In particular, we focus on the two-point correlation function in redshift space, its distortions on small and large scales, and the galaxy pairwise peculiar velocities. Where possible we compare our predictions with the 2dFGRS and we discuss the sensitivity of these clustering statistics to several details regarding the halo occupation statistics. We show that the halo occupation obtained analytically can reliably be implemented in N-body simulations. We find that the standard ΛCDM model, together with the halo occupation we have obtained, can reproduce many of the observational results. However, we find a significant discrepancy between the model predictions and observations on small scales. We show that to get consistent results on small scales, either the mass-to-light ratios for clusters of galaxies are significantly higher than normally assumed, or the linear power spectrum has an amplitude that is significantly lower than its 'concordance' value.

This paper is organized as follows. In Section 2 we review the CLF formalism developed in Papers I and II. Section 3 introduces the N-body simulations and describes our method of populating dark matter haloes in these simulations with galaxies of different type and luminosity. Section 4 investigates several clustering statistics in real space and focuses on the accuracy with which mock galaxy distributions can be constructed using our CLF formalism. In Section 5 we use these mock galaxy distributions to construct mock galaxy redshift surveys that are comparable in size with the 2dFGRS. We extract the redshift-space two-point correlation function from this mock redshift survey, investigate its anisotropies induced by the galaxy peculiar motions, and compare our results to those obtained from the 2dFGRS by Hawkings et al. (2003). In Section 6 we discuss possible ways to alleviate the discrepancy between model and observations on small scales, and we summarize our results in Section 7.

2 THE CONDITIONAL LUMINOSITY FUNCTION

In Paper I we developed a formalism, based on the conditional luminosity function $\Phi(L \mid M)$, to link the distribution of galaxies to that of dark matter haloes. We introduced a parametrized form for $\Phi(L \mid M)$ which we constrained using the LF and the correlation lengths as a function of luminosity. In Paper II we extended this formalism by constructing separate CLFs for early- and late-type galaxies. In this paper we use these results to populate dark matter haloes, obtained from large numerical simulations, with both early- and late-type galaxies of different luminosities. For completeness,

we briefly summarize here the main ingredients of the CLF formalism, and refer the reader to Papers I and II for more details.

The conditional luminosity function is parametrized by a Schechter function:

$$\Phi(L|M)\,\mathrm{d}L = \frac{\tilde{\Phi}^*}{\tilde{L}^*}\left(\frac{L}{\tilde{L}^*}\right)^{\tilde{\alpha}} \exp(-L/\tilde{L}^*)\,\mathrm{d}L, \tag{1}$$

where $\tilde{L}^* = \tilde{L}^*(M)$, $\tilde{\alpha} = \tilde{\alpha}(M)$ and $\tilde{\Phi}^* = \tilde{\Phi}^*(M)$ are all functions of halo mass M.[1] Following Papers I and II, we write the average total mass-to-light ratio of a halo of mass M as

$$\left\langle \frac{M}{L} \right\rangle (M) = \frac{1}{2}\left(\frac{M}{L}\right)_0 \left[\left(\frac{M}{M_1}\right)^{-\gamma_1} + \left(\frac{M}{M_1}\right)^{\gamma_2}\right], \tag{2}$$

which has four free parameters: a characteristic mass M_1, for which the mass-to-light ratio is equal to $(M/L)_0$, and two slopes, γ_1 and γ_2, that specify the behaviour of $\langle M/L\rangle$ at the low- and high-mass ends, respectively. A similar parametrization is adopted for the characteristic luminosity $\tilde{L}^*(M)$:

$$\frac{M}{\tilde{L}^*(M)} = \frac{1}{2}\left(\frac{M}{L}\right)_0 f(\tilde{\alpha}) \left[\left(\frac{M}{M_1}\right)^{-\gamma_1} + \left(\frac{M}{M_2}\right)^{\gamma_3}\right], \tag{3}$$

with

$$f(\tilde{\alpha}) = \frac{\Gamma(\tilde{\alpha}+2)}{\Gamma(\tilde{\alpha}+1,1)}. \tag{4}$$

Here $\Gamma(x)$ is the gamma function and $\Gamma(a,x)$ the incomplete gamma function. This parametrization has two additional free parameters: a characteristic mass M_2 and a power-law slope γ_3. For $\tilde{\alpha}(M)$ we adopt a simple linear function of $\log(M)$:

$$\tilde{\alpha}(M) = \alpha_{15} + \eta \log(M_{15}), \tag{5}$$

with M_{15} the halo mass in units of $10^{15}\,h^{-1}\,\mathrm{M}_\odot$, $\alpha_{15} = \tilde{\alpha}(M_{15} = 1)$, and η describes the change of the faint-end slope $\tilde{\alpha}$ with halo mass. Note that once $\tilde{\alpha}(M)$ and $\tilde{L}^*(M)$ are given, the normalization of the conditional LF, $\tilde{\Phi}^*(M)$, is obtained through equations (1) and (2), using the fact that the total (average) luminosity in a halo of mass M is

$$\langle L\rangle(M) = \int_0^\infty \Phi(L|M) L\,\mathrm{d}L = \tilde{\Phi}^*\tilde{L}^*\Gamma(\tilde{\alpha}+2). \tag{6}$$

Finally, we introduce the mass-scale $M_{\rm min}$ below which we set the CLF to zero; i.e. we assume that no stars form inside haloes with $M < M_{\rm min}$. Motivated by reionization considerations (see Paper I for details) we adopt $M_{\rm min} = 10^9\,h^{-1}\,\mathrm{M}_\odot$ throughout.

In order to split the galaxy population into early and late types, we follow Paper II and introduce the function $f_{\rm late}(L, M)$, which specifies the fraction of galaxies with luminosity L in haloes of mass M that are late type. The CLFs of late- and early-type galaxies are then given by

$$\Phi_{\rm late}(L\,|\,M)\,\mathrm{d}L = f_{\rm late}(L, M)\Phi(L\,|\,M)\,\mathrm{d}L \tag{7}$$

and

$$\Phi_{\rm early}(L\,|\,M)\,\mathrm{d}L = [1 - f_{\rm late}(L, M)]\,\Phi(L\,|\,M)\,\mathrm{d}L. \tag{8}$$

As with the CLF for the entire population of galaxies, $\Phi_{\rm late}(L\,|\,M)$ and $\Phi_{\rm early}(L\,|\,M)$ are constrained by 2dFGRS measurements of the LFs and the correlation lengths as a function of luminosity. We assume that $f_{\rm late}(L, M)$ has a quasi-separable form

$$f_{\rm late}(L, M) = g(L)h(M)q(L, M). \tag{9}$$

[1] Halo masses are defined as the masses within the radius r_{180} inside of which the average overdensity is 180.

Here

$$q(L, M) = \begin{cases} 1 & \text{if} \quad g(L)h(M) \leqslant 1 \\ \dfrac{1}{g(L)h(M)} & \text{if} \quad g(L)h(M) > 1 \end{cases} \tag{10}$$

is to ensure that $f_{\rm late}(L, M) \leqslant 1$. We adopt

$$g(L) = \frac{\hat{\Phi}_{\rm late}(L)}{\hat{\Phi}(L)} \frac{\int_0^\infty \Phi(L\,|\,M)n(M)\,\mathrm{d}M}{\int_0^\infty \Phi(L\,|\,M)h(M)n(M)\,\mathrm{d}M} \tag{11}$$

where $n(M)$ is the halo mass function (Sheth & Tormen 1999; Sheth, Mo & Tormen 2001), $\hat{\Phi}_{\rm late}(L)$ and $\hat{\Phi}(L)$ correspond to the *observed* LFs of the late-type and entire galaxy samples, respectively, and

$$h(M) = \max\left\{0, \min\left[1, \left(\frac{\log(M/M_a)}{\log(M_b/M_a)}\right)\right]\right\} \tag{12}$$

with M_a and M_b two additional free parameters, defined as the masses at which $h(M)$ takes on the values 0 and 1, respectively. As shown in Paper II, this parametrization allows the population of galaxies to be split into early and late types such that their respective LFs and clustering properties are well fitted.

In Papers I and II we presented a number of different CLFs for different cosmologies and different assumptions regarding the free parameters. In what follows we focus on the flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $h = H_0/(100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}) = 0.7$ and with initial density fluctuations described by a scale-invariant power spectrum with normalization $\sigma_8 = 0.9$. These cosmological parameters are in good agreement with a wide range of observations, including the recent *Wilkinson Microwave Anisotropy Probe* (*WMAP*) results (Spergel et al. 2003), and in what follows we refer to it as the 'concordance' cosmology. Finally, we adopt the CLF with the following parameters: $M_1 = 10^{10.94}\,h^{-1}\,\mathrm{M}_\odot$, $M_2 = 10^{12.04}\,h^{-1}\,\mathrm{M}_\odot$, $M_a = 10^{17.26}\,h^{-1}\,\mathrm{M}_\odot$, $M_b = 10^{10.86}\,h^{-1}\,\mathrm{M}_\odot$, $(M/L)_0 = 124\,h\,(\mathrm{M/L})_\odot$, $\gamma_1 = 2.02$, $\gamma_2 = 0.30$, $\gamma_3 = 0.72$, $\eta = -0.22$ and $\alpha_{15} = -1.10$. This model (referred to as model D in Paper II) yields excellent fits to the observed LFs and the observed correlation lengths as a function of both luminosity and type.[2]

3 POPULATING HALOES WITH GALAXIES

3.1 Numerical simulations

The main goal of this paper is to use the CLF described in the previous section to construct mock galaxy redshift surveys, and to study a number of statistical properties of these distributions that can be compared with observations from existing or forthcoming redshift surveys. The distribution of dark matter haloes is obtained from a set of large N-body simulations (dark matter only). The set consists of a total of six simulations with $N = 512^3$ particles each, that

[2] Note that the parameters listed here are slightly different from those given in the original version of Paper II, as they are based on a corrected version of the galaxy luminosity function. As shown in Paper I, a change in the overall amplitude of the luminosity function in the fitting has some effect on the best-fitting values of the correlation lengths. This is due to the combination of the following two effects. First, our model assumes a fixed mass-to-light ratio for massive haloes and so a change in the amplitude of the luminosity function leads to a change in the relative number of galaxies in small/large haloes. Second, although the correlation length as a function of luminosity was used as input in our fitting of the conditional luminosity function, there is some freedom for the 'best-fitting' values of the correlation lengths to change in the fitting, because the error bars on the observed correlation lengths are quite large.

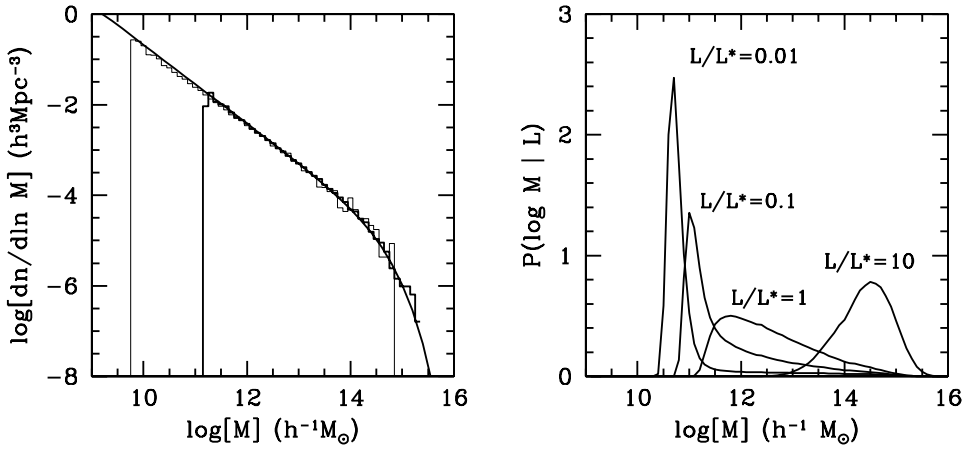

Figure 1. The left-hand panel plots the halo mass functions of the numerical simulations discussed in the text (histograms). The mass function with a low mass cut at about $2 \times 10^{11}\ h^{-1}\ \mathrm{M}_\odot$ corresponds to a simulation with $L_{\rm box} = 300\ h^{-1}$ Mpc, while the other corresponds to a L_{100} simulation with $L_{\rm box} = 100\ h^{-1}$ Mpc. The solid curve is the Sheth et al. (2001) mass function which is shown for comparison. Note the excellent agreement, both between the two simulations and between the simulation results and the theoretical prediction. The right-hand panel plots the conditional probability distributions $P(M\,|\,L)$ for four different luminosities as indicated. $L^* = 1.1 \times 10^{10}\ h^{-2}\ \mathrm{L}_\odot$ is the characteristic luminosity of the Schechter fit to the 2dFGRS LF of Madgwick et al. (2002). Combining these conditional probability distributions with the halo mass functions shown in the left-hand panel gives an indication of the completeness level that can be obtained with both the L_{100} and L_{300} simulations (see text).

have been carried out on the VPP5000 Fujitsu supercomputer of the National Astronomical Observatory of Japan with the vectorized-parallel P^{3}M code (Jing & Suto 2002). Each simulation evolves the distribution of the dark matter from an initial redshift of $z = 72$ down to $z = 0$ in a ΛCDM 'concordance' cosmology. All simulations consider boxes with periodic boundary conditions; in two cases $L_{\rm box} = 100\ h^{-1}$ Mpc while the other four simulations all have $L_{\rm box} = 300\ h^{-1}$ Mpc. Different simulations with the same box size are completely independent realizations and are used to estimate errors due to cosmic variance. The particle masses are $6.2 \times 10^8\ h^{-1}\ \mathrm{M}_\odot$ and $1.7 \times 10^{10}\ h^{-1}\ \mathrm{M}_\odot$ for the small- and large-box simulations, respectively. One of the simulations with $L_{\rm box} = 100\ h^{-1}$ Mpc has previously been used by Jing & Suto (2002) to derive a triaxial model for density profiles of CDM haloes, and we refer the reader to that paper for complementary information about the simulations. In what follows we refer to simulations with $L_{\rm box} = 100\ h^{-1}$ Mpc and $L_{\rm box} = 300\ h^{-1}$ Mpc as L_{100} and L_{300} simulations, respectively.

Dark matter haloes are identified using the standard friends-of-friends (FOF) algorithm (Davis et al. 1985) with a linking length of 0.2 times the mean interparticle separation. Haloes obtained with this linking length have a mean overdensity of ∼180 (Porciani, Dekel & Hoffman 2002), in good agreement with the definition of halo masses used in our CLF analysis. For each individual simulation we construct a catalogue of haloes with 10 particles or more, for which we store the mass (number of particles), the position of the most bound particle, and the mean velocity of the halo and velocity dispersion. Note that the FOF algorithm can sometimes select poor systems (those with a small number of particles) that are spurious and have abnormally large velocity dispersions. We have therefore made a check to make sure that the particles assigned to a system according to the FOF algorithm are gravitationally bound. Our test showed that this correction is important only for low-mass haloes, and that it has almost no effect on our results. The left panel of Fig. 1 plots the $z = 0$ halo mass functions for one of the L_{100} simulations and for one of the L_{300} simulations (histograms), with all spurious haloes erased. For comparison, we also plot (solid line) the analytical halo mass function given in Sheth & Tormen (1999) and Sheth et al. (2001).[3] The agreement is remarkably good, both between the two simulations and between the simulation results and the theoretical prediction.

Note that our choice for box sizes of $100\ h^{-1}$ Mpc and $300\ h^{-1}$ Mpc is a compromise between high mass resolution and a sufficiently large volume to study the large-scale structure. The impact of mass resolution is apparent from considering the conditional probability function

$$P(M\,|\,L)\,\mathrm{d}M = \frac{\Phi(L\,|\,M)}{\Phi(L)} n(M)\,\mathrm{d}M, \tag{13}$$

(see Paper I), which gives the probability that a galaxy of luminosity L resides in a halo with mass in the range $M \pm \mathrm{d}M/2$. The right panel of Fig. 1 plots this probability distribution obtained from the CLF given in Section 2 for four different luminosities: $L = L^*/100$, $L = L^*/10$, $L = L^*$ and $L = 10L^*$. Whereas $10\,L^*$ galaxies are typically found in haloes with $10^{13}\ h^{-1} \lesssim M \lesssim 10^{15}\ h^{-1}\ \mathrm{M}_\odot$, galaxies with $L = L^*/100 \sim 10^8\ h^{-2}\ \mathrm{L}_\odot$ typically reside in haloes of $M \simeq 5 \times 10^{10}\ h^{-1}\ \mathrm{M}_\odot$. Comparing these probability distributions with the halo mass functions in the left panel, we see that the L_{300} simulations can only yield a complete galaxy distribution down to $L \sim 0.4L^*$. The L_{100} simulation, however, resolves dark matter haloes down to masses of $10^{10}\ h^{-1}\ \mathrm{M}_\odot$, which is sufficient to model the galaxy population down to $L \sim 0.01L^*$. On the other hand, luminous galaxies may be underrepresented in this small-box simulation, because it contains fewer massive haloes than expected. Combining these two sets of simulations, however, will enable us to study the clustering properties of galaxies covering a sufficiently large volume and a sufficiently large range of luminosities.

3.2 Halo occupation numbers

When populating haloes with galaxies based on the CLF one first needs to choose a minimum luminosity. Based on the mass resolution of the simulations we adopt $L_{\rm min} = 10^7\ h^{-2}\ \mathrm{L}_\odot$ throughout. The mean occupation number of galaxies with $L \geqslant L_{\rm min}$ for a halo

[3] This same mass function is used in the CLF analysis described in Section 2.

with mass M then follows from the CLF according to:

$$\langle N(M)\rangle = \int_{L_{\rm min}}^{\infty} \Phi(L\,|\,M)\,{\rm d}L. \quad (14)$$

In order to Monte Carlo sample occupation numbers for individual haloes one requires the full probability distribution $P(N\,|\,M)$ (with N an integer) of which $\langle N(M)\rangle$ gives the mean, i.e.

$$\langle N(M)\rangle = \sum_{N=0}^{\infty} N\,P(N\,|\,M). \quad (15)$$

As a simple model we adopt

$$P(N\,|\,M) = \begin{cases} N' + 1 - \langle N(M)\rangle & \text{if} \quad N = N' \\ \langle N(M)\rangle - N' & \text{if} \quad N = N' + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

Here N' is the largest integer smaller than $\langle N(M)\rangle$. Thus, the actual number of galaxies in a halo of mass M is either N' or $N' + 1$. This particular model for the distribution of halo occupation numbers is supported by semi-analytical models and hydrodynamical simulations of structure formation (Benson et al. 2000; Berlind et al. 2003) which indicate that the halo occupation probability distribution is narrower than a Poisson distribution with the same mean. In addition, distribution (16) is successful in yielding power-law correlation functions, much more so than, for example, a Poisson distribution (Benson et al. 2000; Berlind & Weinberg 2002).

3.3 Assigning galaxies their luminosity and type

Since the CLF only gives the average number of galaxies with luminosities in the range $L \pm {\rm d}L/2$ in a halo of mass M, there are many different ways in which one can assign luminosities to the N_i galaxies of halo i, and yet be consistent with the CLF. The simplest approach would be simply to draw N_i luminosities (with $L > L_{\rm min}$) randomly from $\Phi(L\,|\,M)$. Alternatively, one could use a more deterministic approach, and, for instance, always demand that the j th brightest galaxy has a luminosity in the range $[L_j, L_{j-1}]$. Here L_j is defined such that a halo has on average j galaxies with $L > L_j$, i.e.

$$\int_{L_j}^{\infty} \Phi(L\,|\,M)\,{\rm d}L = j. \quad (17)$$

We adopt an intermediate approach in most of our discussion, giving special treatment only to the one brightest galaxy per halo. The luminosity of this so-called 'central' galaxy, $L_{\rm c}$, is drawn from $\Phi(L|\,M)$ with the restriction $L > L_1$ and thus has an expectation value of

$$\langle L_{\rm c}(M)\rangle = \int_{L_1}^{\infty} \Phi(L\,|\,M) L\,{\rm d}L = \tilde{\Phi}^* \tilde{L}^* \Gamma\left(\tilde{\alpha}+2, L_1/\tilde{L}^*\right). \quad (18)$$

The remaining $N_i - 1$ galaxies are referred to as 'satellite' galaxies and are assigned luminosities in the range $L_{\rm min} < L < L_1$, again drawn from the distribution function $\Phi(L\,|\,M)$. In Section 4.2, we test the effect of luminosity sampling by comparing the results obtained from all the three approaches.

Finally, the galaxies are assigned morphological types as follows. For each galaxy with luminosity L in a halo of mass M we draw a random number $\mathcal{R}$ in the range [0, 1]. If $\mathcal{R} < f_{\rm late}(L, M)$ then the galaxy is a late type, otherwise an early type.

3.4 Assigning galaxies their phase-space coordinates

Once the population of galaxies has been assigned luminosities and types, they need to be assigned a position within their halo as well as a peculiar velocity. The central galaxy is assumed to be located at the 'centre' of the corresponding dark halo, which we associate with the position of the most-bound particle, and its peculiar velocity is set equal to the mean halo velocity (cf. Yoshikawa, Jing & Börner 2003). For the satellite galaxies we follow two different approaches. In the first, we assign the $N_i - 1$ satellites the positions and peculiar velocities of $N_i - 1$ randomly selected dark matter particles that are part of the FOF halo under consideration. This thus corresponds to a scenario in which satellite galaxies are completely unbiased with respect to the density and velocity distribution of dark matter particles in FOF haloes. We refer to satellite galaxies populated this way as 'FOF satellites'.

We also consider a more analytical model for the satellite distribution. This allows us first of all to assess whether a simple analytical description can be found to describe the population of satellite galaxies, and secondly, provides us with a simple framework to investigate the sensitivity of various clustering statistics to details regarding the density and velocity bias of satellite galaxies. We assume that the number density distribution of satellite galaxies follows a NFW density distribution (Navarro, Frenk & White 1997):

$$\rho(r) = \frac{\bar{\delta}\bar{\rho}}{(r/r_{\rm s})(1+r/r_{\rm s})^2}, \quad (19)$$

where $r_{\rm s}$ is a characteristic radius, $\bar{\rho}$ is the average density of the Universe, and $\bar{\delta}$ is a dimensionless amplitude which can be expressed in terms of the halo concentration parameter $c = r_{180}/r_{\rm s}$ as

$$\bar{\delta} = \frac{180}{3}\frac{c^3}{\ln(1+c) - c/(1+c)}. \quad (20)$$

Here r_{180} is the radius inside of which the halo has an average overdensity of 180. Numerical simulations show that halo concentration depends on halo mass, and we use the relation given by Bullock et al. (2001), converted to the c appropriate for our definition of halo mass. The radial number density distribution of satellite galaxies is assumed to follow equation (19) with a concentration $c_{\rm g} = c$, and the angular position is assumed to be random over the 4π solid angle. Peculiar velocities are assumed to be the sum of the peculiar (mean) velocity of the host halo plus a random velocity which is assumed to be distributed isotropically and to follow a Gaussian, one-dimensional velocity distribution:

$$f(v_j) = \frac{1}{\sqrt{2\pi}\sigma_{\rm gal}} \exp\left(-\frac{v_j^2}{2\sigma_{\rm gal}^2}\right). \quad (21)$$

Here v_j is the velocity relative to that of the central galaxy along the j-axis, and $\sigma_{\rm gal}$ is the one-dimensional velocity dispersion of the galaxies, which we set equal to that of the dark matter particles, $\sigma_{\rm dm}$, in the halo under consideration. We refer to satellite galaxies populated this way as 'NFW satellites'.

4 RESULTS IN REAL SPACE

Figs 2 and 3 show slices of mock galaxy distributions (hereafter MGDs) constructed from L_{100} and L_{300} simulations, respectively. Satellite galaxies are assigned positions and velocities using the NFW scheme outlined above. Results are shown for all galaxies (upper right panels), and separately for early types (lower-right panels) and late types (lower-left panels). For comparison, we also show the distribution of dark matter particles in the upper-left panels. Note how the large-scale structure in the dark matter distribution is delineated by the distribution of galaxies, and that early-type galaxies are more strongly clustered than late-type galaxies.

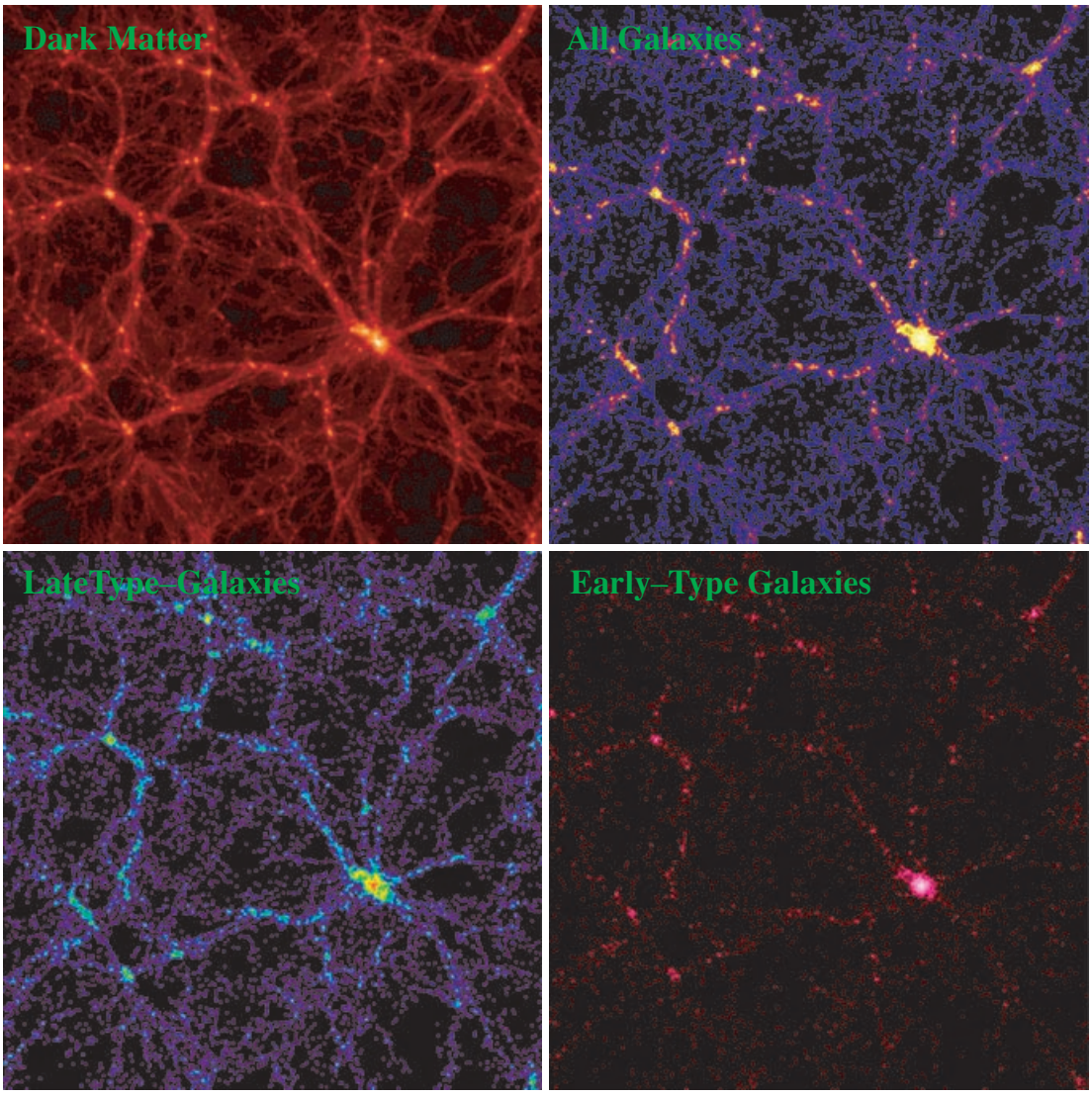

Figure 2. Projected dark matter/galaxy distributions of a $100 \times 100 \times 10\ h^{-1}$ Mpc slice in one of the L_{100} mock galaxy distributions. The panels show (clockwise from top-left) the dark matter particles, all galaxies (early plus late), early-type galaxies, and late-type galaxies. Galaxies are weighted by their luminosities. Note how the galaxies trace the large-scale structure of the dark matter, and how early-type galaxies are more strongly clustered than late-type galaxies.

In this section we discuss the general, *real-space* properties of these MGDs. In Section 5 below we construct mock galaxy redshift surveys to investigate the impact of redshift distortions. The main goal of this section, however, is to investigate with what accuracy the combination of numerical simulations and our CLF analysis can be used to construct self-consistent mock galaxy distributions. In particular, we want to examine to what accuracy these MGDs can recover the input used to constrain the CLFs. Note that this is not a trivial question. The CLF modelling is based on the halo model, which only yields an approximate description of the dark matter distribution in the non-linear regime (see the discussions in Cooray & Sheth 2002; Huffenberger & Seljak 2003). In addition, as described in Section 3, the CLF alone does not yield sufficient information to construct MGDs, and we had to make additional assumptions regarding the distribution of galaxies within individual haloes. A further goal of this section, therefore, is to investigate how these assumptions impact on the clustering statistics.

4.1 The luminosity function

The CLFs used to construct the MGDs shown in Fig. 2 and 3 are constrained by the 2dFGRS luminosity functions for early- and late-type galaxies obtained by Madgwick et al. (2002). Therefore, as long as the halo mass function is well sampled by the simulations, the LFs of our MGDs should match those of Madgwick et al. (2002). Fig. 4 shows a comparison between the 2dFGRS LFs (symbols with error bars) and those recovered from the MGDs (solid lines). To emphasize the level of agreement between the recovered LFs and the input LFs, Fig. 5 plots the ratio between the two. Over a large range of luminosities, the recovered LFs match the observational input extremely well. In the L_{300} simulation, however, the LFs are underestimated for $L \lesssim 3 \times 10^9\ h^{-2}\ \mathrm{L}_\odot$ ($M_{b_J} - 5 \log h \gtrsim -18.4$). This is due to the absence of haloes with $M \lesssim 2 \times 10^{11}\ h^{-1}\ \mathrm{M}_\odot$ (see Fig. 1). Note how this discrepancy sets in at higher L for the late-type galaxies than for the early-types, because the latter are preferentially located in more massive haloes. For the early-types the L_{300} mock

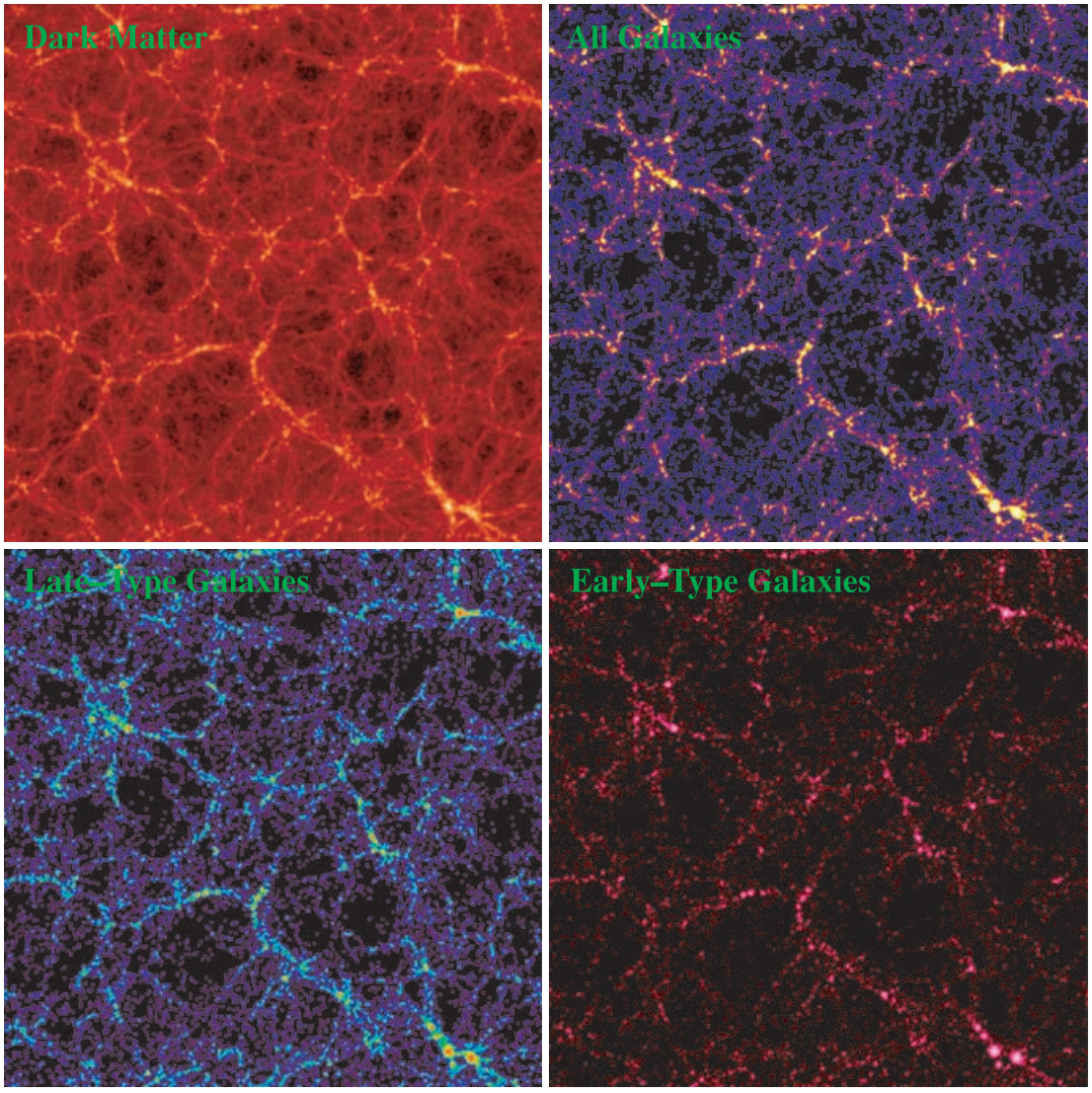

Figure 3. Same as Fig. 2, but for a $300 \times 300 \times 20\,h^{-1}$ Mpc slice taken from one of the L_{300} mock galaxy distributions.

is virtually complete down to $M_{b_J} - 5 \log h \;\; -17$ (see fig. 10 of Paper II), reflecting the fact that only a very small fraction of the early-type galaxies brighter than this magnitude reside in haloes below the mass resolution limit. In the L_{100} simulations, on the other hand, the LFs accurately match the data down to the faintest luminosities, but here the MGD underestimates the LFs at the bright end ($M_{b_J} - 5 \log h \lesssim -22$). This is due to the limited box size, which causes the number of massive haloes (the main hosts of the brightest galaxies) to be underestimated (cf. Fig. 1). Note that even the LFs of the L_{300} simulations underestimate the *observed* number of bright galaxies. This reflects a small inaccuracy of our CLF to match the observed bright end of the LFs accurately (see Paper II).

4.2 The real-space correlation function

In addition to the LFs of early- and late-type galaxies, the CLFs used here to construct our MGDs are also constrained by the luminosity and type dependence of the correlation lengths as measured from the 2dFGRS by Norberg et al. (2002a). Here we check to what degree this 'input' is recovered from the MGDs.

The left panel of Fig. 6 plots the real-space two-point correlation functions (2PCFs) for dark matter particles in the L_{100} (dashed line) and L_{300} (dotted line) simulations. The solid line corresponds to the evolved, non-linear dark matter correlation function of Smith et al. (2003) and is shown for comparison.[4] As one can see, on large scales ($r \gtrsim 6\ h^{-1}$ Mpc) the correlation amplitude obtained from the L_{100} simulations is systematically lower than both that obtained from the L_{300} simulations and that obtained from the fitting formula of Smith et al., suggesting that the box-size effect is non-negligible in the L_{100} simulations. Note also that the large-scale correlation amplitude given by the L_{300} simulations is slightly higher than the model of Smith et al. It is unclear if this discrepancy is due to the inaccuracy of the fitting formula, or due to cosmic variance in the present simulations. As we will see below, this discrepancy limits the accuracy of model predictions.

The right-hand panel of Fig. 6 plots the 2PCFs for the galaxies in the L_{100} (dashed line) and L_{300} (dotted line) MGDs. Note how

[4] In fitting the CLF we have used this function to compute the correlation length of the dark matter (see Paper II).

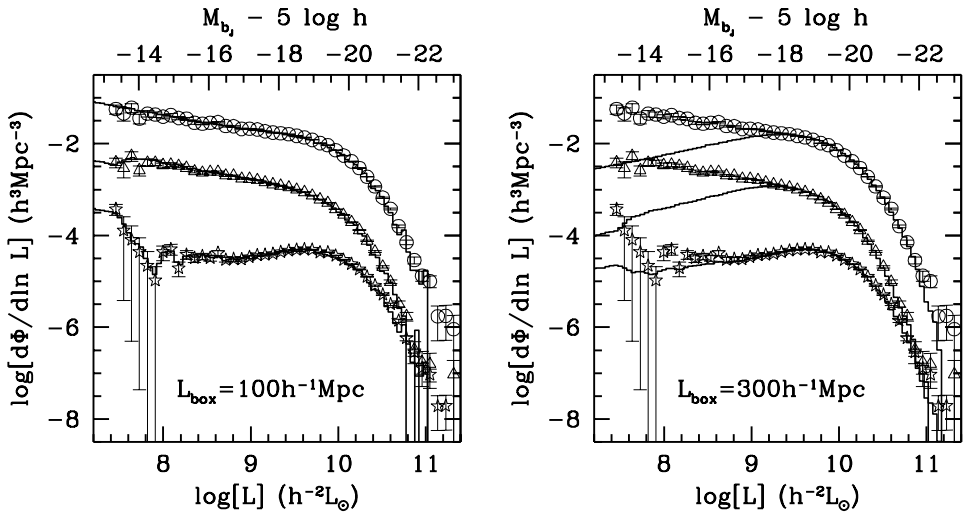

Figure 4. The luminosity functions of the mock galaxies constructed from the L_{100} (left) and L_{300} (right) halo catalogues (solid lines). For comparison, we also plot the LFs obtained by Madgwick et al. (2002) for all galaxies (circles), for late-type galaxies (triangles) and for early-type galaxies (stars). For clarity, the latter two LFs have been shifted down by one and two orders of magnitude in the y-direction, respectively. Except for incompleteness effects due to the sampling of the halo mass function (see text for details), the mock galaxy distributions have LFs that are in excellent agreement with the data.

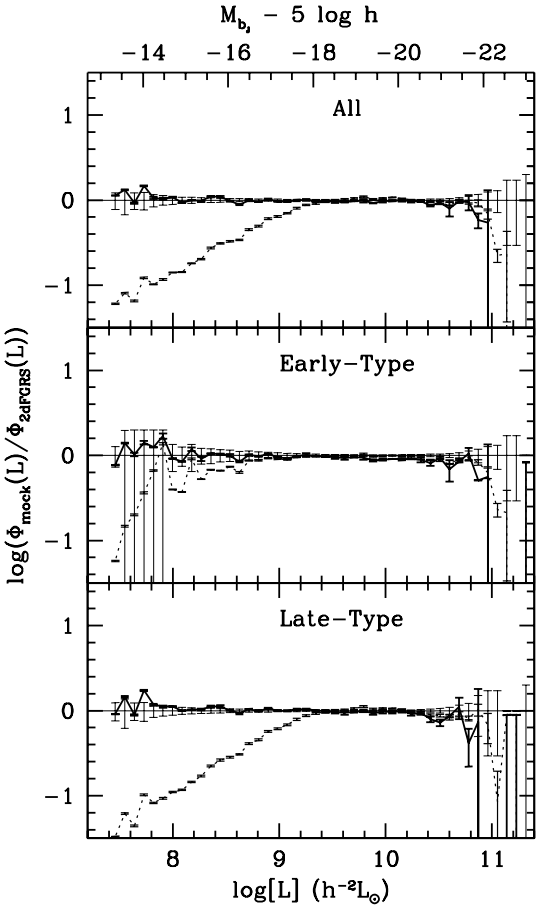

Figure 5. The ratio of the luminosity function of mock galaxies, $\Phi_{\rm mock}(L)$, to that of the 2dFGRS, $\Phi_{\rm 2dFGRS}(L)$ (taken from Madgwick et al. 2002). The thin error bars indicate the errors on $\Phi_{\rm 2dFGRS}(L)$. The thick solid (dashed) lines correspond to the LFs obtained from the L_{100} (L_{300}) simulations. The error bars for the mock galaxies are obtained from the 1σ variance of the two L_{100} and the four L_{300} simulations, respectively. See text for discussion.

the galaxies reveal the same trend on large scales as the dark matter particles, with larger correlations in the L_{300} than in the L_{100} MGD.

Fig. 7 shows the correlation lengths r_0 as a function of luminosity for all (upper panel), early-type (middle panel) and late-type (lower panel) galaxies. These have been obtained by fitting $\xi(r)$ with a power-law relation of the form $\xi(r) = (r/r_0)^{-\gamma}$ over the same range of scales as used by Norberg et al. (2002a). Solid squares and open stars correspond to correlation lengths obtained from the L_{300} and L_{100} MGDs, respectively. Note that the error bars on the predicted correlation lengths are based on the scatter among independent simulations boxes. They are significantly smaller than the error bars on the observational data, because the model predictions are based on real-space correlation functions, while the observational results are based on projected correlation functions in redshift space. The agreement with the data (open circles) is reasonable, even though several systematic trends are apparent. In particular, the correlation lengths obtained from the L_{300} simulation are slightly higher than the observations while the opposite applies to the L_{100} simulation. These discrepancies are due to two effects. First of all, as shown in Fig. 6 the dark matter on large scales is more strongly clustered in the L_{300} simulations than in the L_{100} simulations. That this can account for most of the differences between the scalelengths obtained from the L_{300} and L_{100} simulations, is illustrated by the dotted and solid horizontal lines, which indicate the correlation lengths of the dark matter particles in the L_{300} and L_{100} simulations, respectively. Secondly, the measured correlation lengths correspond to a non-zero, median redshift which is larger for the more luminous galaxies. In determining the best-fitting parameters for the CLF this redshift effect is taken into account (see Papers I and II). However, in the construction of our MGDs, we only use the dark matter distribution at $z = 0$. As discussed in Paper I, this can overestimate the correlation length by about 10 per cent. Given these sources of systematic errors, one should be careful not to overinterpret any discrepancy between the correlation lengths in the mock survey and those obtained from real redshift distributions.

In order to investigate the sensitivity of the 2PCF in the MGDs to the way we assign luminosities and phase-space coordinates to the

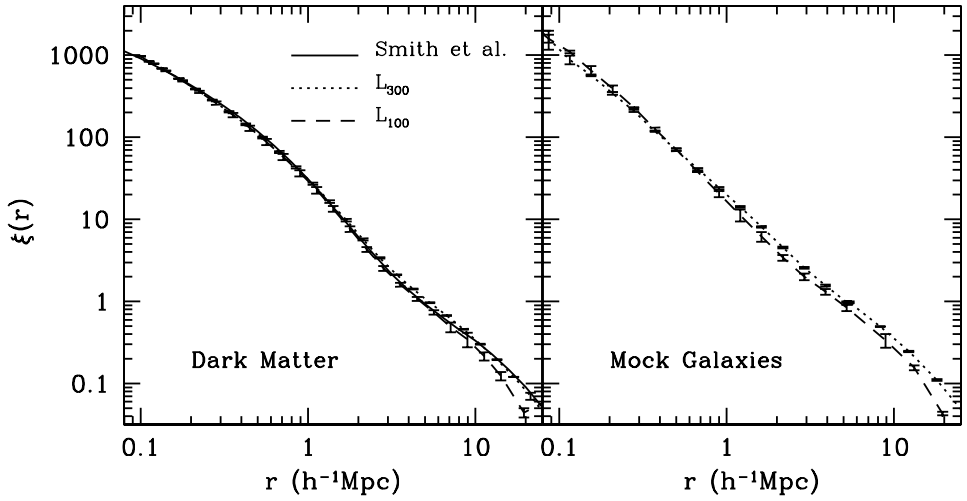

Figure 6. Two-point correlation functions for dark matter particles (left panel) and mock galaxies (right panel). The dotted and dashed lines correspond to results from the L_{300} and L_{100} simulations, respectively. The solid line in the left panel corresponds to the evolved, non-linear correlation function for the dark matter obtained by Smith et al. (2003), and is shown for comparison. Due to the limited box sizes, the L_{300} (L_{100}) simulations slightly over (under) predict the correlation power on large scales with respect to the model of Smith et al. The 2PCFs in the right panel are calculated for galaxies with absolute magnitudes $M_{b_J} - 5\log h < -18.4$, which corresponds to the completeness limit of the L_{300} MGDs. Note that the box size also affects the 2PCFs of the mock galaxies on large scales. Error bars are the variance among the two (L_{100}) and four (L_{300}) independent realizations.

galaxies within the dark matter haloes, we construct MGDs using one of the L_{300} simulations with different models for the luminosity assignment and spatial distribution of satellite galaxies within haloes. We have confirmed that using one of the L_{100} simulations instead yields the same results. We first test the impact of the luminosity assignment. Here, instead of the fiducial model for the luminosity assignment (the intermediate approach discussed in Section 3.3), we use both the deterministic and random assignments (see Section 3.3 for definitions) to construct the MGDs. In Fig. 8 we shown the ratios between the correlation functions obtained from these MGDs and those obtained from the fiducial MGD. For bright galaxies, the deterministic model gives the lowest amplitudes on small scales ($r \lesssim 1\,h^{-1}$ Mpc), while the random model gives the highest amplitudes. This is expected. The mean number of bright galaxies in a typical halo is not much larger than 1 and so not many close pairs of bright galaxies are expected in the deterministic model. More such pairs are expected in the random model because more than one galaxy in a typical halo can be assigned a large luminosity due to random fluctuations. The dashed lines in Fig. 8 correspond to a MGD with FOF satellites (see Section 3.4). The agreement of the 2PCFs between this MGD with FOF satellites and our fiducial MGD indicate that the spherical NFW model is a good approximation of the average density distribution of dark matter haloes. We have also tested the impact of changing the concentration of galaxies, $c_{\rm g}$; increasing (decreasing) $c_{\rm g}$ with respect to the dark matter halo concentration, c, increases (decreases) the 2PCFs on small scales ($r \lesssim 1\,h^{-1}$ Mpc). However, even when changing the ratio $c_{\rm g}/c$ by a factor of 2, the amplitude of this change is smaller than the differences resulting from changing the luminosity assignment.

All in all, changes in the way we assign luminosities and phase-space coordinates to the galaxies only have a mild impact on the 2PCFs, and only at small scales $\lesssim 1\,h^{-1}$ Mpc. This is in good agreement with Berlind & Weinberg (2002) who have shown that these effects are much smaller than changes in the second moment of the halo occupation distributions. For example, assuming a Poissonian $P(N\,|\,M)$, rather than equation (16) has a much larger impact on the 2PCFs than any of the changes investigated above. As we show in Section 5 below, with the $P(N\,|\,M)$ of equation (16) we obtain correlation functions that are in better agreement with observations, providing empirical support for this particular occupation number distribution.

It is interesting to note that although small changes in the way we assign luminosities and phase-space coordinate do not have a big impact on the statistical measurements we are considering here, such changes can lead to quite different results for other statistical measures. As shown in van den Bosch et al. (in preparation), various statistics of satellite galaxies around bright galaxies can be used to distinguish models that make similar predictions about the clustering on large scales.

4.3 Pairwise velocities

The peculiar velocities of galaxies are determined by the action of the gravitational field, and so are directly related to the matter distribution in the Universe. Observationally, the properties of galaxy peculiar velocities are inferred from distortions in the correlation function. We defer this discussion to Section 5. Here we derive statistical quantities directly from the simulated peculiar velocities of galaxies.

We define the pairwise peculiar velocity of a galaxy pair as

$$v_{12}(r) \equiv [\boldsymbol{v}(\boldsymbol{x}+\boldsymbol{r}) - \boldsymbol{v}(\boldsymbol{x})] \cdot \hat{\boldsymbol{r}}, \tag{22}$$

with $\boldsymbol{v}(\boldsymbol{x})$ the peculiar velocity of a galaxy at $\boldsymbol{x}$. The mean pairwise peculiar velocity and the pairwise peculiar velocity dispersion (PVD) are

$$\langle v_{12}(r)\rangle \quad \text{and} \quad \sigma_{12}(r) \equiv \left\langle [v_{12}(r) - \langle v_{12}(r)\rangle]^2 \right\rangle^{1/2}, \tag{23}$$

where $\langle\cdots\rangle$ denotes an average over all pairs of separation r.

In order to gain insight, we compute $\langle v_{12}(r)\rangle$ and $\sigma_{12}(r)$ from the L_{300} simulations for both dark matter particles and for galaxies with $M_{b_J} - 5\log h \leqslant -18.4$ (which corresponds to the completeness limit of these simulations, see Fig. 4).

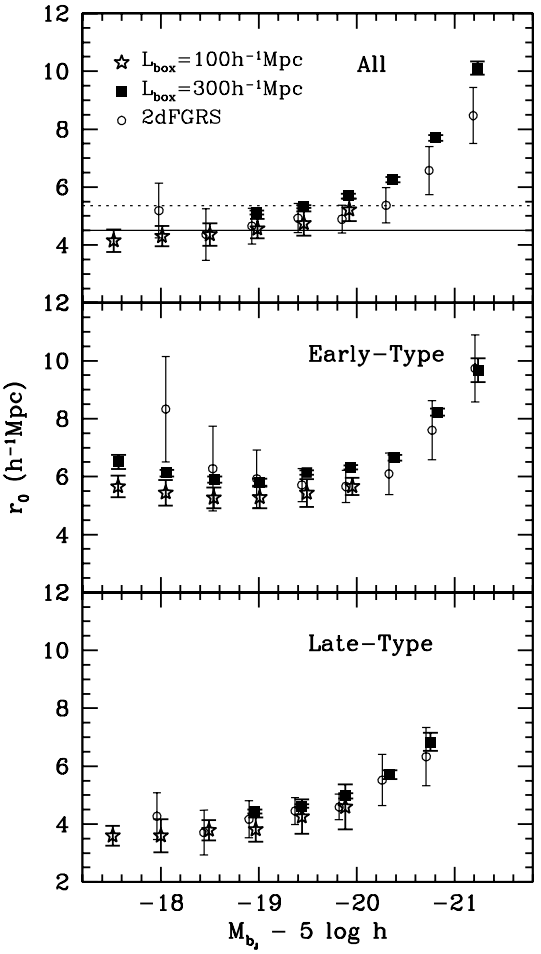

Figure 7. The real-space correlation length, r_0, as a function of galaxy luminosity and type. The top panel shows the results for the combined sample of early- plus late-type galaxies, while the middle (bottom) panel shows results for the early- (late-) type galaxies only. Solid squares and stars correspond to the correlation lengths obtained from the L_{300} and L_{100} simulations, respectively. The error bars correspond to the 1σ variance from the two (four) independent realizations for L_{100} (L_{300}). We also indicate (open circles with error bars) the correlation lengths obtained from the 2dFGRS by Norberg et al. (2002a). In the upper panel, we also plot the correlation lengths for dark matter particles for L_{100} (solid line) and L_{300} (dotted line) simulations. Although the agreement between data and MGDs is reasonable there are small but significant differences. The reason for these discrepancies is discussed in the text.

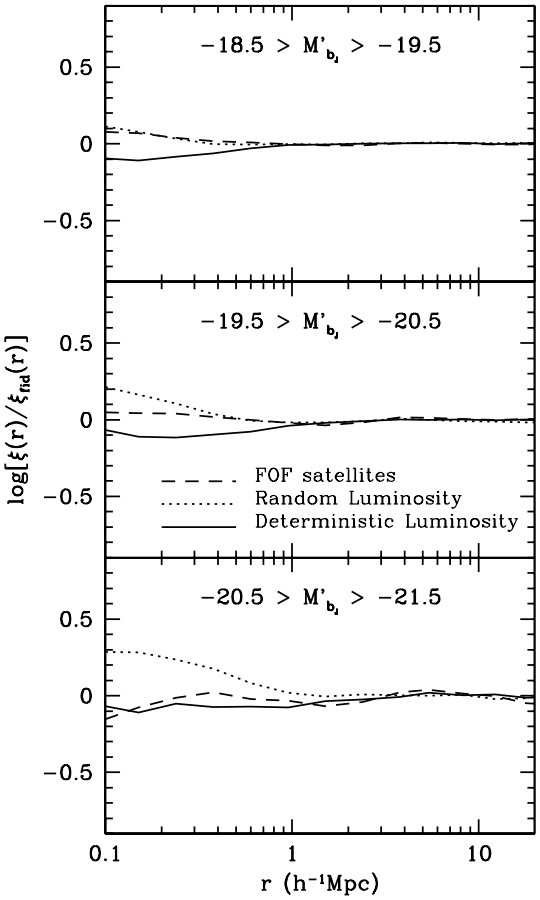

Figure 8. The ratio of the 2PCF $\xi(r)$ in three MGDs compared to that of our fiducial MGD. The only difference among these various MGDs is the way that we assign luminosities and phase-space coordinates to the galaxies. Solid (dotted) lines correspond to a MGD in which we use a deterministic (random) method to assign galaxies their luminosities (see Section 3.3 for definitions). In the MGD corresponding to the dashed line we use the intermediate, fiducial method to assign luminosities, but here we use 'FOF satellites' rather than 'NFW satellites' (see Section 3.4 for definitions). Results are shown for galaxies in three different magnitude bins (as indicated) in one of the L_{300} simulations. However, results for the L_{100} simulations look virtually identical.

Results are shown in Fig. 9. The upper-left panel compares the mean pairwise peculiar velocities of the dark matter particles (solid circles) with those of two realizations of the galaxies: one with 'NFW satellites' (open circles) and the other with 'FOF satellites' (stars). At sufficiently small separations, one probes the virialized regions of dark matter haloes, and one thus finds $\langle v_{12}\rangle = 0$. At larger separations, one starts to probe the infall regions around the virialized haloes, yielding negative values for $\langle v_{12}(r)\rangle$. Finally, at sufficiently large separations $\langle v_{12}(r)\rangle \to 0$ due to the large-scale homogeneity and isotropy of the Universe.

Both the dark matter particles and the galaxies from our MGDs indeed reveal such behaviour, with $\langle v_{12}(r)\rangle$ peaking at $\sim 3\,h^{-1}$ Mpc. However, there is a markedly strong difference between the $\langle v_{12}(r)\rangle$ of galaxies in the MGD with NFW satellites and that of the dark matter. In this particular MGD, the galaxies experience significantly smaller infall velocities than the dark matter particles. However, this difference between dark matter and galaxies is almost absent in the MGD with FOF satellites. This is due to the fact that in the NFW model, we populate satellites with isotropic velocity dispersions within a sphere of radius r_{180}. We are thus assuming that the entire region out to r_{180} is virialized in that there is no net infall. However, simple collapse models predict that for our concordance cosmology only the region out to r_{340} (i.e. the radius inside of which the average overdensity is 340) is virialized (Bryan & Norman 1998). The difference between the MGDs with NFW satellites and FOF satellites indicates that the regions between r_{340} and r_{180} are still infalling, resulting in non-zero $\langle v_{12}\rangle$.

In the lower-left panel, we compare the PVDs for galaxies and dark matter particles. Here the MGDs with FOF satellites and NFW satellites are fairly similar, and significantly lower than for the dark matter. This can be understood as follows. At small separations, the PVD is a *pair*-weighted measure for the potential well in which dark matter particles (galaxies) reside. For the galaxies in our MGDs the halo occupation number per unit mass, N/M, decreases with the

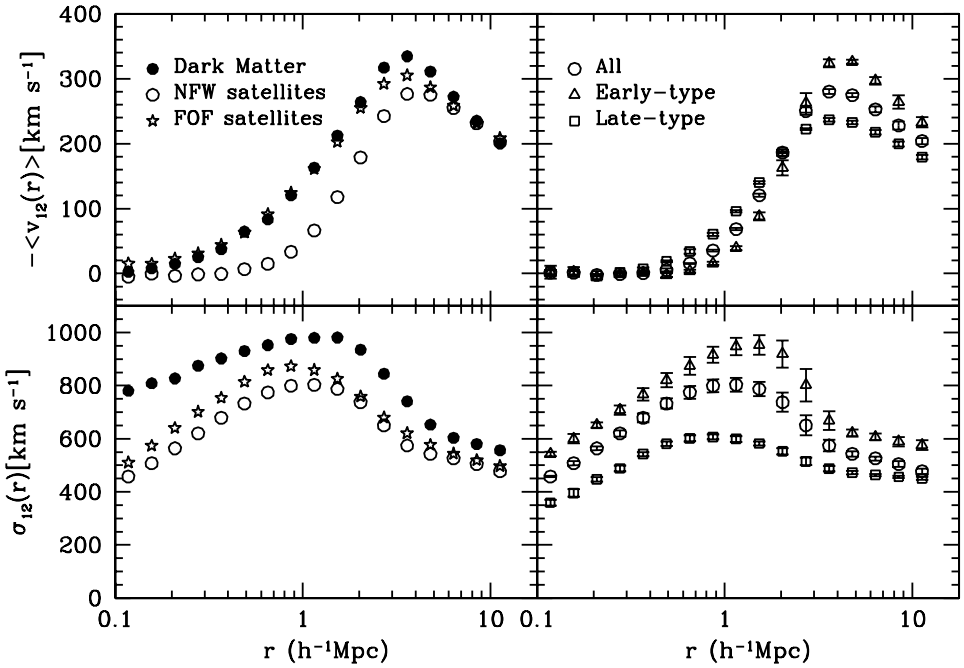

Figure 9. The mean pairwise velocities (upper panels) and pairwise velocity dispersions (lower panels) estimated from the three-dimensional (real-space) velocities of the mock galaxies and dark matter particles. All results correspond to the L_{300} simulations only. Left-hand panels compare dark matter particles (solid circles) with galaxies either with NFW satellites (open circles) or with FOF satellites (open stars). Right-hand panels display the galaxy-type dependence for a model with NFW satellites (error bars indicate the rms scatter among the four independent L_{300} simulations). See text for detailed discussion.

mass of dark matter haloes (see Paper II). Therefore, the massive haloes (with larger velocity dispersions) contribute relatively less to the PVDs of galaxies. Although the difference between the $\sigma_{12}(r)$ of the MGDs with FOF and NFW satellites shows that the PVDs have some dependence on the details regarding the infall regions around virialized haloes, these effects are typically small.

The upper-right and lower-right panels of Fig. 9 show how $\langle v_{12}(r)\rangle$ and $\sigma_{12}(r)$ depend on galaxy type. Results are shown for the MGD based on NFW satellites. The mean velocities for early-type galaxies are larger than those for late-type galaxies on large scales, but smaller on small scales. In addition, the PVD of early-type galaxies is higher than that of late-type galaxies on all scales. All these differences are easily understood as a reflection of the fact that early-type galaxies are preferentially located in the larger, more massive haloes which have larger velocity dispersions and larger infall velocities.

Fig. 10 shows the pairwise velocity distributions for four different separations r, within a logarithmic interval of $\Delta\log r = 0.125$. On small scales, the distribution is well-fitted by an exponential for both dark matter particles and galaxies. This validates the assumption made in earlier analyses about this distribution (Davis & Peebles 1983; Mo, Jing & Börner 1993; Fisher et al. 1994; Marzke et al. 1995). It is also consistent with earlier results obtained from theoretical models and numerical simulations based on dark matter particles (Efstathiou et al. 1988; Diaferio & Geller 1996; Sheth 1996; Mo, Jing & Börner 1997; Seto & Yokoyama 1998; Magira, Jing & Suto 2000). For larger separations $f(v_{12})$ is skewed towards negative values of v_{12}, because galaxies tend to approach each other due to gravitational infall. Clearly, a single exponential function is no longer a good approximation to the pairwise peculiar velocity distribution at large separations. Although for $v_{12} < 0$ (infall) the exponential remains remarkably accurate, for $v_{12} > 0$ the pairwise velocity distribution reveals a more Gaussian behaviour. This may have important implications for the derivation of PVDs (especially at large separations), which typically is based on the assumption of a purely exponential $f(v_{12})$. We shall return to this issue in more detail in Section 5.2.

5 RESULTS IN REDSHIFT SPACE

The statistical quantities of galaxy clustering discussed in the previous section are based on real distances between galaxies in our MGDs. However, because of the peculiar velocities of galaxies, such quantities cannot be obtained directly from a galaxy redshift survey. On small scales the virialized motion of galaxies within dark matter haloes cause a reduction of the correlation power, while on larger scales the correlations are boosted due to the infall motion of galaxies towards overdensity regions (Kaiser 1987; Hamilton 1992). As discussed in the introduction, these distortions contain useful information about the universal density parameter, the bias of galaxies on large (linear) scales, and the pairwise velocities of galaxies.

In this section, we use the MGDs presented above to construct large mock galaxy redshift surveys (hereafter MGRSs). The main goals are to compare various clustering statistics from these mock surveys with observational data from the 2dFGRS, and to investigate how the details about the CLF and the distribution of galaxies within haloes impact on these statistics. For the model–data comparison we use the large-scale structure analysis of Hawkins et al. (2003, hereafter H03), which is based on a subsample of the 2dFGRS consisting of all galaxies located in the North Galactic Pole (NGP) and South Galactic Pole (SGP) survey strips with redshift $0.01 \leqslant z \leqslant 0.20$ and apparent magnitude $b_{\rm J} < 19.3$. This sample consists of $\sim$166 000 galaxies covering an area on the sky of $\sim$1090 deg^2.

In order to carry out a proper comparison between model and observation, we aim to construct MGRSs that have the same selections as the 2dFGRS. First of all, the survey depth of $z_{\rm max} = 0.2$ implies

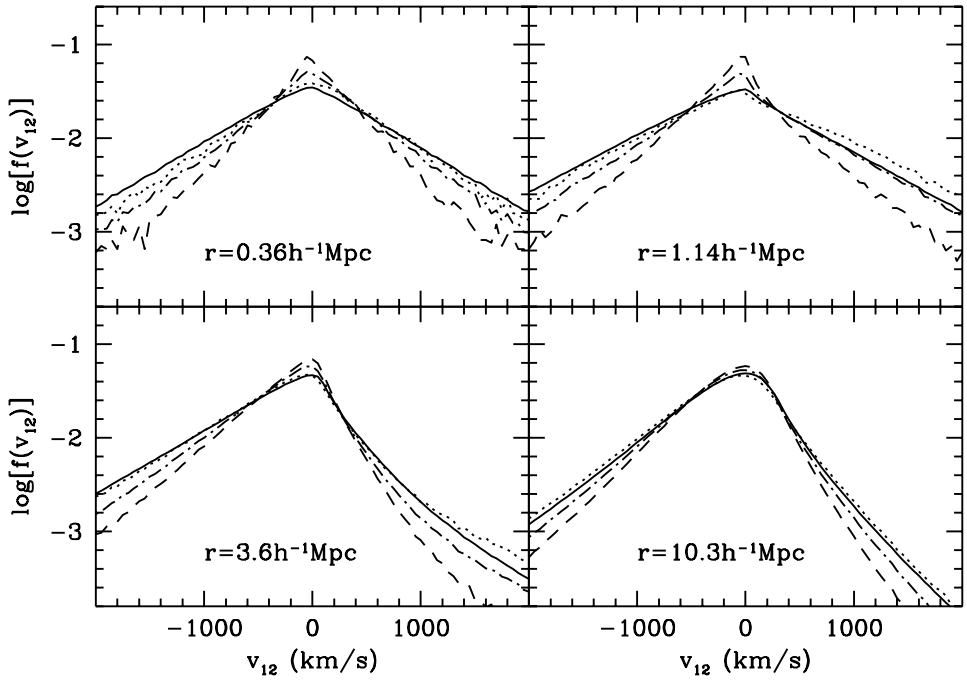

Figure 10. Distribution of pairwise velocities, $f(v_{12})$, for dark matter particles (solid curves), and for mock galaxies in the L_{300} simulation. Results are shown for four separations r as indicated, and for all galaxies (dot-dashed lines), early-type galaxies (dotted lines) and late-type galaxies (dashed lines). On small scales ($r < 1\ h^{-1}$ Mpc) the pairwise velocity distributions are symmetric and reveal an obvious exponential form. On larger scales, however, $f(v_{12})$ reveals clear asymmetries: for $v_{12} < 0$ the distribution is still close to an exponential, while for $v_{12} > 0$ the distribution more resembles a normal distribution.

that we need to cover a volume with a depth of 600 h^{-1} Mpc, i.e. twice that of our big L_{300} simulations. In principle, we could stack $4 \times 4 \times 4$ identical L_{300} boxes (which have periodic boundary conditions), so that a depth of 600 h^{-1} Mpc can be achieved in all directions for an observer located at the centre of the stack. However, there is one problem with this set-up; as we have shown in Figs 1 and 4 the L_{300} MGD is only complete down to $M_{b_J} - 5\log h \simeq -18.4$. Taking account of the apparent magnitude limit of the survey, this implies that our MGRSs would be incomplete out to a distance of $\sim$350 h^{-1} Mpc. We can overcome this problem by using the higher-resolution L_{100} simulation, which is complete down to $M_{b_J} - 5\log h \simeq -14$. We therefore replace the central $2 \times 2 \times 2$ L_{300} boxes with a stack of $6 \times 6 \times 6$ L_{100} boxes. The final lay-out of our virtual universe is illustrated in Fig. 11. Unless stated otherwise, satellite galaxies are assigned to dark matter haloes based on our standard NFW method described in Section 3.4.

Observational selection effects, which are modelled according to the final public data release of the 2dFGRS (see also Norberg et al. 2002b), are taken into account using the following steps.

(i) We place a virtual observer at the centre of the stack of boxes (the solid dot in Fig. 11), define a (α, δ)-coordinate frame, and remove all galaxies that are not located in the areas equivalent to the NGP and SGP regions of the 2dFGRS.

(ii) Next, for each galaxy we compute the redshift as 'seen' by the virtual observer. We take the observational velocity uncertainties into account by adding a random velocity drawn from a Gaussian distribution with dispersion 85 km s^{-1} (Colless et al. 2001).

(iii) We compute the apparent magnitude of each galaxy according to its luminosity and distance. Since galaxies in the 2dFGRS were pruned by apparent magnitude before a k-correction was applied, we proceed as follows. We first apply a negative k-correction, then select galaxies according to the position-dependent magnitude limit (obtained using the apparent magnitude limit masks provided by the 2dFGRS team), and finally k-correct the magnitudes back to their rest-frame b_J-band. Throughout we use the type-dependent k-corrections given in Madgwick et al. (2002).

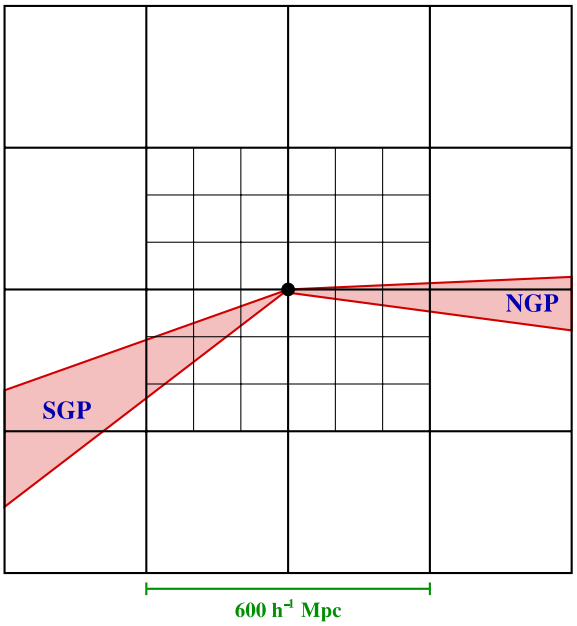

Figure 11. The stacking geometry of the L_{100} and L_{300} simulation boxes used to construct the MSB mock galaxy redshift surveys. The virtual observer is located at the centre of the stack, indicated by a thick solid dot. Note that for MGRSs in the MB set, the stack of $6 \times 6 \times 6$ L_{100} boxes is replaced by a stack of $2 \times 2 \times 2$ L_{300} boxes.

(iv) To mimic the position- and magnitude-dependent completeness of the 2dFGRS, we randomly sample each galaxy using the completeness masks provided by the 2dFGRS team. The incompleteness of the 2dFGRS parent sample is taking into account by randomly discarding 9 per cent of all mock galaxies (Norberg et al. 2002b).

(v) Finally, we mimic the actual selection criteria of the 2dFGRS sample used in H03 by restricting the sample to galaxies within the redshift range $0.01 \leqslant z \leqslant 0.20$ and with completeness $\geqslant 0.7$.

Each MGRS thus constructed contains, on average, 169 000 galaxies, with a dispersion of ∼5000 due to cosmic variance. The number of galaxies in our mock catalogues are consistent with the observations at the 1σ level. Note that the correlation functions presented by H03 have been corrected for the observational bias due to fibre collisions, and we therefore do not mimic these effects in our MGRSs.

Since we have two L_{100} simulations and four L_{300} simulations, we construct $2 \times 4 = 8$ mock catalogues with different combinations of small- and big-box simulations. In what follows, we refer to this set of mock catalogues as MSBs (for Mock Small/Big). As an example, Fig. 12 shows the distribution of a subset of galaxies in one of these mock catalogues. Although each of our MSB catalogues covers an extremely large volume, and thus should not be very sensitive to cosmic variance, it is constructed using simulations with box sizes of 100 and 300 h^{-1} Mpc only. If, for instance, the L_{100} simulation contains a big cluster, the $6 \times 6 \times 6$ reproduction of this box in our MGRSs might introduce some unrealistic features. Furthermore, as shown in Section 4.2, the L_{100} box underestimates the amount of clustering on large scales. Therefore, this set of MGRSs, which replicates this box 27 times, might underestimate the clustering on large scales as well. In order to test the sensitivity of our results to these potential problems, and to have a better handle on the impact of cosmic variance in our mock surveys, we construct four alternative MGRSs. Each consists of a $4 \times 4 \times 4$ stack of one of the four L_{300} simulations (i.e. we replace the $6 \times 6 \times 6$ stack of L_{100} boxes by a $2 \times 2 \times 2$ stack of L_{300} boxes). In what follows we refer to this set of mock catalogues as MBs (for Mock Big). These MGRSs, although incomplete for $M_{b_J} - 5 \log h > -18.4$, should not suffer from the lack of clustering power on large scales. The MSB set, on the other hand, does not suffer from incompleteness, but instead lacks some large-scale power. As we will see below, both the MSB and MB mocks give similar results on large scales, suggesting that the box-size effect does not have a significant influence on our results.

5.1 Two-point correlation functions

From our MGRSs we compute $\xi(r_p, \pi)$ using the estimator (Hamilton 1993)

$$\xi(r_p, \pi) = \frac{\langle RR \rangle \langle DD \rangle}{\langle DR \rangle^2} - 1 \tag{24}$$

with $\langle DD \rangle$, $\langle RR \rangle$, and $\langle DR \rangle$ the number of galaxy–galaxy, random–random, and galaxy–random pairs with separation (r_p, π). Here r_p and π are the pair separations perpendicular and parallel to the line-of-sight, respectively. Explicitly, for a pair $(\boldsymbol{s}_1, \boldsymbol{s}_2)$, with $\boldsymbol{s}_i = cz_i \hat{\boldsymbol{r}}_i / H_0$, we define

$$\pi = \frac{\boldsymbol{s} \cdot \boldsymbol{l}}{|\boldsymbol{l}|}, \qquad r_p = \sqrt{\boldsymbol{s} \cdot \boldsymbol{s} - \pi^2} \tag{25}$$

Here $\boldsymbol{l} = \frac{1}{2}(\boldsymbol{s}_1 + \boldsymbol{s}_2)$ is the line-of-sight intersecting the pair, and $\boldsymbol{s} = \boldsymbol{s}_1 - \boldsymbol{s}_2$. Random samples are constructed using two different methods. The first uses the mean galaxy number density at redshift z calculated from the 2dFGRS LF. The second randomizes the coordinates of all mock galaxies within the simulation box. Both methods yield indistinguishable estimates of $\xi(r_p, \pi)$ and in what follows we only use the former. Following H03 each galaxy in a pair with redshift separation s is weighted by the factor

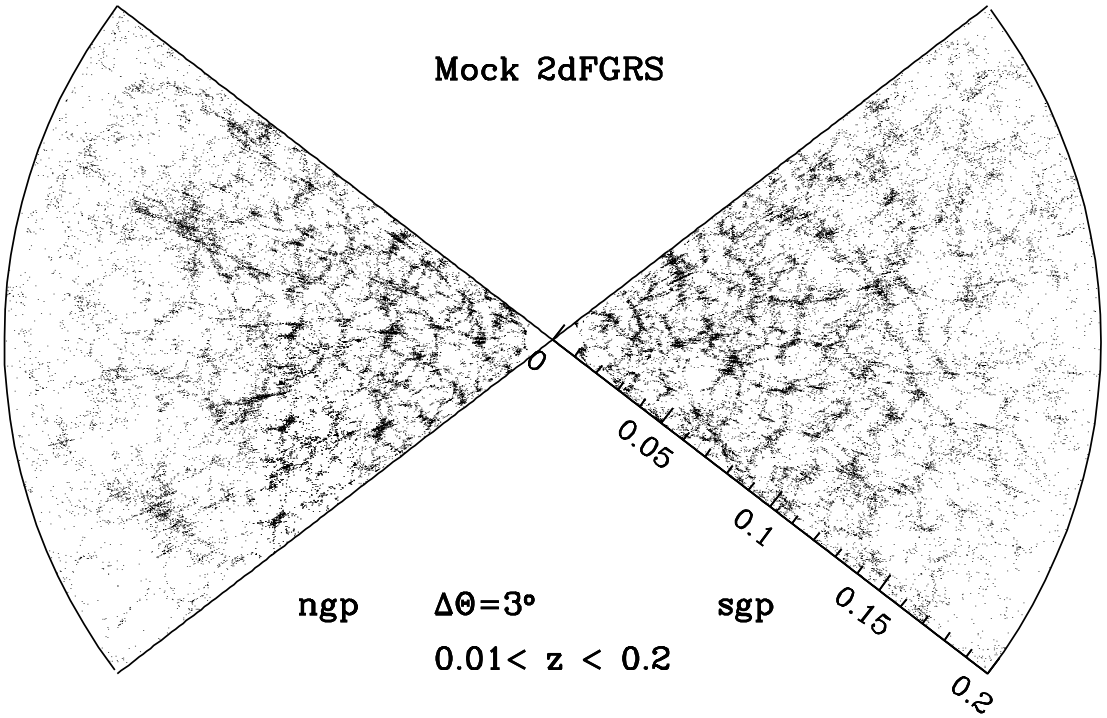

Figure 12. The distribution of a subset of galaxies in one of the MSB mock samples. For clarity, we plot galaxies only in two 3° slices, one in the 'North Galactic Pole' region (NGP) and the other in the 'South Galactic Pole' region (SGP). Only galaxies with redshifts in the range $0.01 < z < 0.2$ are plotted.

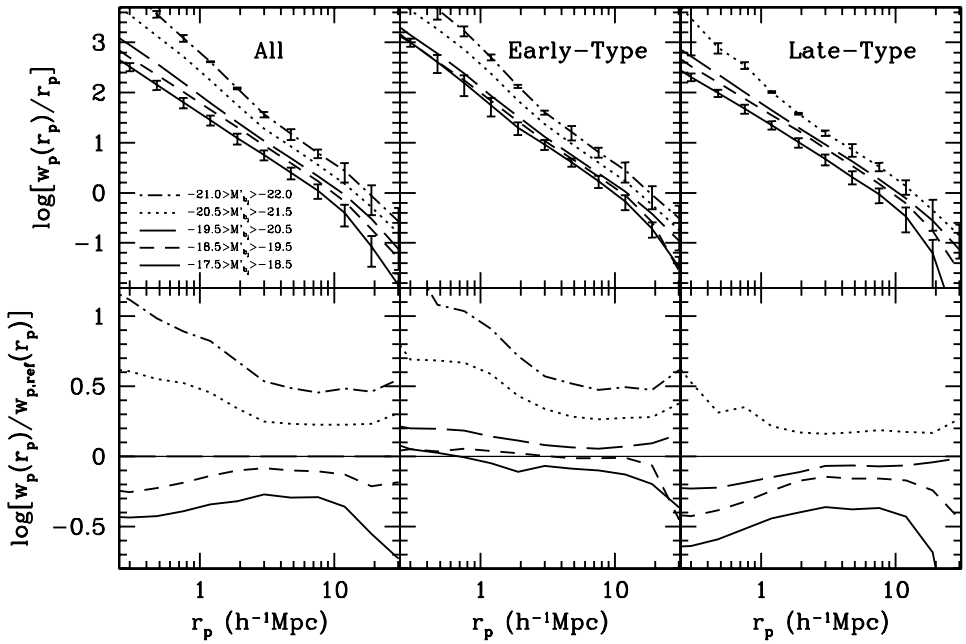

Figure 13. The upper panels show the projected 2PCFs $w_p(r_p)/r_p$ for galaxies of different luminosity and type. The error bars correspond to the 1σ variance among distinct MGRSs (i.e. among the eight MSBs for the faintest subsamples, and among the four MBs for the brightest subsamples). For clarity, the error bars are only plotted for the brightest and faintest subsamples. The lower panels plot the ratios of these $w_p(r_p)$ to that of a reference sample. The reference sample contains all galaxies within the magnitude range $-19.5 > M'_{b_J} > -20.5$ (with $M'_{b_J} = M_{b_J} - 5\log h$). Note that the faintest subsamples, which are impacted by the box-size effect of the L_{100} simulation, reveal a 'break' at $r_p \simeq 10\,h^{-1}$ Mpc.

$$w_i = \frac{1}{1 + 4\pi n(z_i) J_3(s)} \tag{26}$$

with $n(z)$ the number density distribution as a function of redshift and $J_3(s) = \int_0^s \xi(s')s'^2\,\mathrm{d}s'$. Hence each galaxy–galaxy, random–random, and galaxy–random pair is given a weight $w_i w_j$. We replace $\xi(s')$ with a power law using the same parameters as in H03. This redshift-dependent weighting scheme is designed to minimize the variance on the estimated correlation function (Davis & Huchra 1982; Hamilton 1993).

Since the redshift-space distortions only affect π, the projection of $\xi(r_p, \pi)$ along the π-axis can get rid of these distortions and give a function that is more closely related to the real-space correlation function. In fact, this projected 2PCF is related to the real-space 2PCF through a simple Abel transform

$$w_p(r_p) = \int_{-\infty}^{\infty} \xi(r_p, \pi)\,\mathrm{d}\pi = 2\int_{r_p}^{\infty} \xi(r)\frac{r\,\mathrm{d}r}{\sqrt{r^2 - r_p^2}} \tag{27}$$

(Davis & Peebles 1983). Therefore, if the real-space 2PCF is a power law, $\xi(r) = (r_0/r)^\gamma$, the projected 2PCF $w(r_p)$ can be written as

$$w_p(r_p) = \sqrt{\pi}\frac{\Gamma(\gamma/2 - 1/2)}{\Gamma(\gamma/2)}\left(\frac{r_0}{r_p}\right)^\gamma r_p. \tag{28}$$

We start our investigation of the redshift-space clustering properties by computing $w_p(r_p)$ for a number of luminosity bins and for early- and late-type galaxies separately. To compare these projected correlation functions with the 2dFGRS results from Norberg et al. (2002a), we estimate $w_p(r_p)$ using volume-limited samples with the same redshift and magnitude selection criteria as those adopted by Norberg et al. (2002a). For the MSB mocks (which use a stack of $6 \times 6 \times 6$ L_{100} boxes), however, these $w_p(r_p)$ reveal a systematic 'break' at $r_p \sim 10\,h^{-1}$ Mpc. As we have shown in Section 4.2, this is due to the fact that, because of the small box size of the L_{100} simulation, the 2PCF is too small on large scales (see Figs 6 and 7). We can circumvent this problem by using MGRSs from the MB set, in which the stack of $6 \times 6 \times 6$ L_{100} boxes is replaced by a stack of $2 \times 2 \times 2$ L_{300} boxes. However, these MGRSs are only complete down to $M_{b_J} - 5\log h \simeq -18.4$ and can therefore only be used for galaxies brighter than this.

The upper panels of Fig. 13 plot $w_p(r_p)$ for different magnitude bins and for early- and late-type galaxies separately. Except for the faintest magnitude bin, these projected correlation functions are obtained from MGRS in the MB set. Results for the magnitude bin with $-17.5 > M_{b_J} - 5\log h > -18.5$ (solid lines) are obtained from the MSB set. As discussed in Paper II, the projection significantly washes out the features in the real-space 2PCFs at $\sim 2\,h^{-1}$ Mpc, and the projected 2PCFs better resemble a power law. The exception is the $w_p(r_p)$ for the faintest subsample of galaxies, where the 'break' mentioned above is clearly visible. To highlight the luminosity and type dependence of $w_p(r_p)$, the lower panels of Fig. 13 plot the ratios of $w_p(r_p)$ to that of a reference sample defined as all (early-type plus late-type) galaxies with $-19.5 > M_{b_J} - 5\log h > -20.5$. For a given luminosity, the correlation amplitude is higher, and the slope is steeper for early-type galaxies than for late-type galaxies. Significant changes in the slope (and thus deviations from a perfect power law) occur at separations $r_p \sim 2\,h^{-1}$ Mpc, which is at least qualitatively in agreement with recent results from the SDSS (Zehavi et al. 2003).

In order to facilitate a more direct comparison with the 2dFGRS data, we fit a single power-law relation of the form (28) to these $w_p(r_p)$ over the range $2\,h^{-1}\,\mathrm{Mpc} < r_p < 15\,h^{-1}$ Mpc. This range is also adopted by Norberg et al. (2002a) when fitting the projected 2PCFs obtained from the 2dFGRS. Fig. 14 plots the real-space correlation lengths r_0 and the slopes γ thus obtained as a function of luminosity and galaxy type. The agreement between our MGRSs and the 2dFGRS is acceptable. The slight but

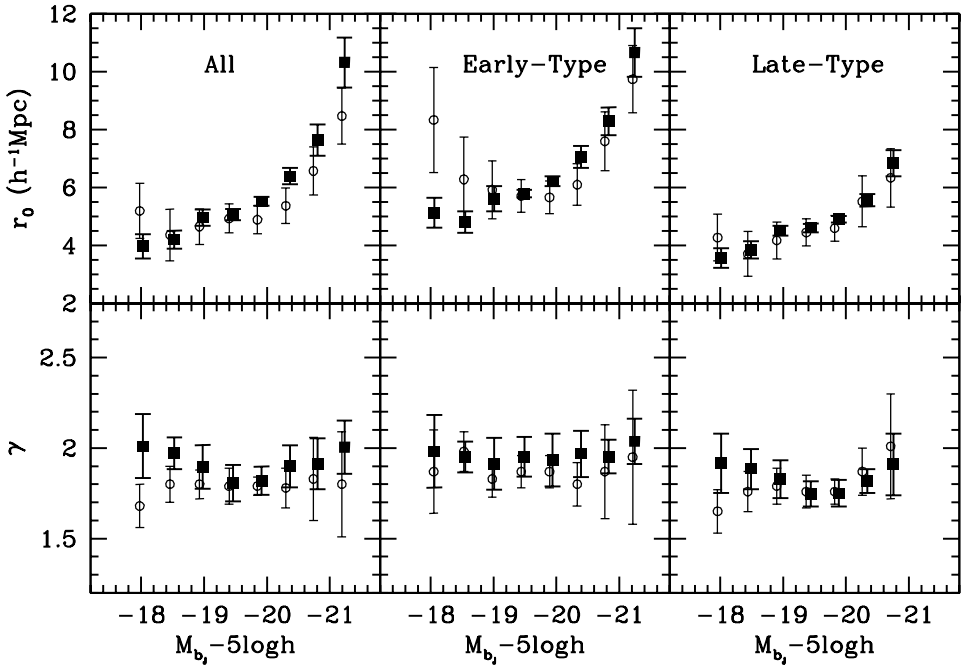

Figure 14. The correlation lengths, r_0, and slopes, γ, of the power laws that best fit the projected correlation functions over the range $2 \leqslant r_p \leqslant 15\,h^{-1}$ Mpc (solid squares). The results for the two faintest luminosity bins are based on the mean and variance of the sample of eight MSB mocks, while results for the other bins are based on the mean and variance of the sample of four MB mocks. Open circles with error bars correspond to the 2dFGRS data of Norberg et al. (2002a), and are shown for comparison. Except for a systematic overestimate of the correlation lengths, the cause of which has been discussed in Section 4.2, there is good agreement between our MGRSs and the 2dFGRS.

systematic overestimate of r_0 is due to the effects discussed in Section 4.2.

We now turn to a comparison of the projected correlation function for the entire, flux limited surveys. The upper-left panel of Fig. 15 compares the $w_p(r_p)$ obtained from our eight MSB and four MB MGRSs with that of the 2dFGRS obtained by H03. The projected correlation functions from our MSBs and MBs agree well with each other (i.e. the 1σ error bars overlap), and, at $r_p \gtrsim 3\,h^{-1}$ Mpc, with the 2dFGRS results. Note that at $r_p \gtrsim 10\,h^{-1}$ Mpc the $w_p(r_p)$ obtained from the MB mocks is slightly larger than that obtained from the MSB mocks, again due to the effects discussed in Section 4.2.

At large scales, $w_p(r_p)$ is predominantly sensitive to the halo occupation numbers $\langle N(M)\rangle$ and virtually independent of the second moment of $P(N\,|\,M)$ or of details regarding the spatial distribution of satellite galaxies. The good agreement at large scales among different MGRSs and with the observations therefore strongly supports our CLF and it shows that any 'cosmic variance' among the different MGRSs has only a relatively small impact on $w_p(r_p)$. On small scales, however, the MGRSs reveal more correlation power (by about a factor 2) than observed. On such scales, $w_p(r_p)$ is sensitive to our assumptions about the second moment of $P(N\,|\,M)$ and, to a lesser degree, the spatial distribution of satellite galaxies. We shall return to this small-scale mismatch and its implications in Section 6 below.

Rather than projecting $\xi(r_p, \pi)$, one may also average $\xi(r_p, \pi)$ along constant $s = \sqrt{r_p^2 + \pi^2}$, yielding the redshift-space 2PCFs $\xi(s)$. The upper-right panel of Fig. 15 plots $\xi(s)$ obtained from our MGRSs, compared to the 2dFGRS results from H03. We find similar behaviour as with the projected correlation function; the eight MSBs and four MBs agree quite well with each other and with the observations at $s \gtrsim 6\,h^{-1}$ Mpc. At smaller redshift-space separations, however, the MGRSs slightly overpredict the correlation power. Note that the MB samples predict higher $\xi(s)$ on small scales than the MSB samples. This difference comes from the fact that the MB samples are incomplete for galaxies fainter than $M_{b_J} - 5\log h = -18.4$. To test this we construct a mock survey from the MSB sample, but only accepting galaxies brighter than this. This yields a $\xi(s)$ in excellent agreement with that of the MB samples over all scales. Thus, although the use of only large-box simulations can result in systematic errors on small scale, the use of small-box simulations in the MSB samples does not cause any significant, systematic errors on large scale.

5.2 Redshift-space distortions

We now turn to a comparison of the detailed shape of $\xi(r_p, \pi)$. In particular, we focus on the distortions with respect to the real-space correlation function $\xi(r)$ induced by the peculiar velocities of galaxies.

The two-dimensional correlation function $\xi(r_p, \pi)$ is often modelled as a convolution of the real-space 2PCF $\xi(r)$ and the conditional distribution function $f(v_{12}\,|\,r)$:

$$1 + \xi(r_p, \pi) = \int_{-\infty}^{\infty} \left[1 + \xi\left(\sqrt{r_p^2 + (\pi - v_{12}/H_0)^2}\right)\right] f(v_{12}\,|\,r)\,\mathrm{d}v_{12} \quad (29)$$

(Peebles 1980). Here v_{12} corresponds to the pairwise peculiar velocity *along the line-of-sight* and r corresponds to the real-space separation. It is standard practice to assume an exponential form for $f(v_{12}|\,r)$ and to ignore its dependence on separation r (cf. Davis & Peebles 1983; Mo et al. 1993, 1997; Fisher et al. 1994; Marzke et al. 1995; Guzzo et al. 1997; Jing et al. 2002; Zehavi et al. 2002). However, as we have shown in Section 4, the exponential form is only

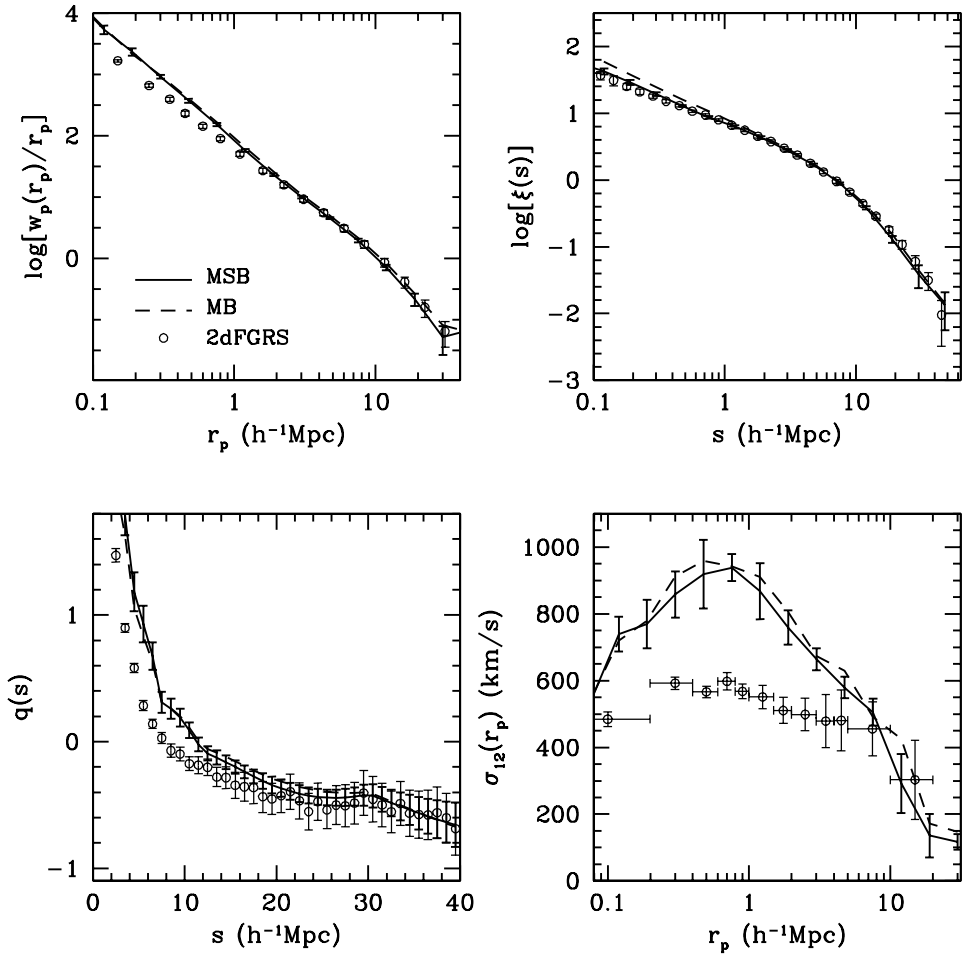

Figure 15. The projected correlation function $w_p(r_p)$ (top-left panel), the redshift-space correlation function $\xi(s)$ (top-right), the quadrupole-to-monopole ratio $q(s)$ (bottom-left), and the PVDs (bottom-right) for the samples of MSB (solid lines) and MB (dashed lines) surveys. Error bars, which are similar for MB and MSB results, are only shown for the MSB results for clarity. These error bars are based on the variance of the eight MSB surveys. The open circles with error bars correspond to the 2dFGRS results obtained by Hawkins et al. (2003), and are shown for comparison. Note that the MSBs and MBs give approximately the same results, but that there are marked differences between model predictions and observations. Note also that the model error bars are in general larger than the difference in the mean between MB and MSB results, implying that these error bars are statistical.

adequate at small separations, and the PVD varies quite strongly with separation. Furthermore, equation (29) is only valid for an isotropic velocity field in the limit where the probability of a real-space pair separation r is independent of the probability of an associated relative velocity v_{12}. Although perhaps a reasonable approximation on small, highly non-linear, scales, it is certainly not valid in linear theory where the velocity and density fields are tightly coupled. In an attempt to partially correct for this, one often assumes that $f(v_{12})$ is the probability distribution for the *relative* velocity about the mean. Using the self-similar infall model, this mean pairwise peculiar velocity, $\langle v_{12}\rangle$, is modelled as

$$\langle v_{12}\rangle(r) = -H_0 F\left(\frac{y}{1+(r/r_0)^2}\right) \tag{30}$$

(Davis & Peebles 1977) with $y = |\pi - v_{12}/H_0|$ the separation in real-space along the line-of-sight. $F = 0$ corresponds to a universe without any flow other than the Hubble expansion, while $F = 1$ corresponds to stable clustering. Given the fairly ad hoc nature of this model, and the strong sensitivity to the uncertain value of F (Davis & Peebles 1983), great care is required when interpreting any results based on this model.

A more robust model is based on linear theory and directly modelling the infall velocities around density perturbations. Following Kaiser (1987) and Hamilton (1992) one can write the observed correlation function on linear scales as

$$\xi_{\rm lin}(r_p,\pi) = \xi_0(s)\mathcal{P}_0(\mu) + \xi_2(s)\mathcal{P}_2(\mu) + \xi_4(s)\mathcal{P}_4(\mu). \tag{31}$$

Here $\mathcal{P}_l(\mu)$ is the lth Legendre polynomial, and μ is the cosine of the angle between the line-of-sight and the redshift-space separation $\boldsymbol{s}$. According to linear perturbation theory the angular moments can be written as

$$\xi_0(s) = \left(1+\frac{2\beta}{3}+\frac{\beta^2}{5}\right)\xi(r), \tag{32}$$

$$\xi_2(s) = \left(\frac{4\beta}{3}+\frac{4\beta^2}{7}\right)\left[\xi(r)-\overline{\xi}(r)\right], \tag{33}$$

$$\xi_4(s) = \frac{8\beta^2}{35}\left[\xi(r)+\frac{5}{2}\overline{\xi}(r)-\frac{7}{2}\hat{\xi}(r)\right], \tag{34}$$

with

$$\overline{\xi}(r) = \frac{3}{r^3}\int_0^r \xi(r')r'^2\,{\rm d}r', \tag{35}$$

and

$$\hat{\xi}(r) = \frac{5}{r^5}\int_0^r \xi(r')r'^4\,\mathrm{d}r'. \quad (36)$$

Given a value for β and the real-space correlation function, which can be obtained from $\xi(r_p, \pi)$ via the projected correlation function $w_p(r_p)$, equation (31) yields a model for $\xi(r_p, \pi)$ *on linear scales* that takes proper account of the coupling between the density and velocity fields. To model the non-linear virialized motions of galaxies within dark matter haloes, one convolves this $\xi_{\rm lin}(r_p, \pi)$ with the distribution function of pairwise peculiar velocities $f(v_{12} \mid r)$.

$$1 + \xi(r_p, \pi) = \int_{-\infty}^{\infty} [1 + \xi_{\rm lin}(r_p, \pi - v_{12}/H_0)] f(v_{12} \mid r)\,\mathrm{d}v_{12}. \quad (37)$$

Thus, by modelling $\xi\ (r_p, \pi)$ one can hope to get both an estimate of β as well as information regarding the pairwise peculiar velocity distribution. We follow H03, and assume that the real-space 2PCF is a pure power law, $\xi(r) = (r/r_0)^{-\gamma}$, and that $f(v_{12} \mid r)$ is an exponential that is independent of the real-space separation r:

$$f(v_{12} \mid r) = f(v_{12}) = \frac{1}{\sqrt{2}\sigma_{12}} \exp\left(-\frac{\sqrt{2}|v_{12}|}{\sigma_{12}}\right). \quad (38)$$

Using a simple χ^2 minimization technique, we fit these models, described by the four parameters β, σ_{12}, r_0, and γ, to the $\xi\ (r_p, \pi)$ in each of our eight MSB and four MB MGRSs. The χ^2 is defined as

$$\chi^2 = \sum \left(\frac{\log[1 + \xi]_{\rm model} - \log[1 + \xi]_{\rm data}}{\log[1 + \xi + \Delta\xi]_{\rm data} - \log[1 + \xi - \Delta\xi]_{\rm data}}\right)^2, \quad (39)$$

where the summation is over the $\xi(r_p, \pi)$ data grid with the restriction $8\ h^{-1}$ Mpc $< s < 20\ h^{-1}$ Mpc (see H03) and $\Delta\xi(r_p, \pi)$ is the rms of $\xi(r_p, \pi)$ determined from each of our eight MSB MGRSs (or of our four MB MGRSs). The averages (over the eight or the four MGRSs) of the best-fitting values for β, σ_{12}, r_0, and γ, along with the variances among different samples, are listed in the first two lines of Table 1. These should be compared with the values listed in the last line, which correspond to the best-fitting values obtained from the 2dFGRS by H03 using exactly the same method. As one can see, the best-fitting values for β and correlation lengths of the MSB and the MB sample agree with each other and with the 2dFGRS value at better than the 1σ level. On the other hand, the discrepancies regarding γ and σ_{12} are significant, both of which are significantly higher in our MGRSs than in the 2dFGRS.

In order to investigate these discrepancies in more detail we compute two statistics of the redshift space distortions which we compare to the 2dFGRS. As above, we take great care in using exactly the same method and assumptions as H03. Therefore, even if some aspects of the model are questionable, this allows a meaningful comparison of our results with those obtained by H03. First of all, we compute the modified quadrupole-to-monopole ratio

$$q(s) \equiv \frac{\xi_2(s)}{\frac{3}{s^3}\int_0^s \xi_0(s')s'^2\,\mathrm{d}s' - \xi_0(s)}. \quad (40)$$

where $\xi_l(s)$ is given by

$$\xi_l(s) = \frac{2l+1}{2}\int_{-1}^{1} \xi(r_p, \pi)\mathcal{P}_l(\mu)\,\mathrm{d}\mu. \quad (41)$$

The lower-left panel of Fig. 15 plots $q(s)$ for MSBs and MBs together with the 2dFGRS results (open circles with error bars). Although the MGRSs reveal the same overall behaviour as the 2dFGRS data, and are mutually consistent, they systematically overpredict $q(s)$. On small scales, where random peculiar velocities cause a rapid increase of $q(s)$, this indicates that the virialized motions in our mock surveys are larger than observed (see also below). On large scales, where $q(s)$ is asymptotic to the linear theory value of

$$q(s) = \frac{-\frac{4}{3}\beta - \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2}, \quad (42)$$

this might indicate that the value of β inherent to our MGRSs is too small compared to the real Universe. On the other hand, Cole, Fisher & Weinberg (1994) have shown that non-linear, small-scale power can affect $q(s)$ out to fairly large separations. Therefore the systematic overestimate of q at large s may simply be a reflection of the random peculiar velocities being too large, rather than an inconsistency regarding the value of β.

The second statistic that we use to compare the redshift-space distortions in our MGRSs with those of the 2dFGRS are the PVDs, $\sigma_{12}(r_p)$, as a function of projected radius, r_p. Following H03, we keep r_0, γ and β fixed at the 'global' values listed in Table 1 and determine $\sigma_{12}(r_p)$ by minimizing χ^2 in a number of independent r_p bins.[5] The results are shown in the lower-right panel of Fig. 15. Whereas the 2dFGRS reveals a $\sigma_{12}(r_p)$ that is almost constant with radius at about 500–600 km s^{-1}, our MGRSs reveal a strong increase from $\sigma_{12} \sim 600$ km s^{-1} at $r_p = 0.1\ h^{-1}$ Mpc to $\sigma_{12} \sim 900$ km s^{-1} at $r_p = 1.0\ h^{-1}$ Mpc, followed by a decrease to $\sigma_{12} \sim 500$ km s^{-1} at $r_p = 10\ h^{-1}$ Mpc. Thus, at around $1\ h^{-1}$ Mpc, our MGRSs dramatically overestimate the PVD. Although there is a non-negligible amount of scatter among the different mock surveys, reflecting the extreme sensitivity of the PVDs to the few richest systems in the survey, the variance among the eight (four) MGRSs is small compared to the discrepancy.

As shown by Peacock et al. (2001) the best-fitting values of σ_{12} and β are highly degenerate. We have tested the impact of this

Table 1. Best-fitting parameters.

Survey (1)	β (2)	r_0 (3)	γ (4)	σ_{12} (5)
MSBs	0.52 ± 0.05	5.78 ± 0.23	1.99 ± 0.03	687 ± 37
MBs	0.56 ± 0.06	5.93 ± 0.27	1.95 ± 0.03	732 ± 33
$(M/L)_{\rm cl}$	0.52 ± 0.08	4.96 ± 0.21	1.88 ± 0.05	487 ± 44
$b_{\rm vel}$	0.47 ± 0.06	5.99 ± 0.19	2.00 ± 0.06	497 ± 47
σ_8	0.51 ± 0.06	5.19 ± 0.17	1.91 ± 0.05	505 ± 25
2dFGRS	0.49 ± 0.09	5.80 ± 0.25	1.78 ± 0.06	514 ± 31

The values of β, r_0 (in h^{-1} Mpc), γ and σ_{12} (in km s^{-1}) that best fit the $\xi(r_p, \pi)$ for $8\ h^{-1}$ Mpc $< s < 20\ h^{-1}$ Mpc for a number of different MGRSs. Note that four MGRSs are used for MBs, while eight MGRSs are used for all other cases. The quoted values are the mean and 1σ variance of these MGRSs. The MGRSs denoted by '$(M/L)_{\rm cl}$' are similar to the MGRSs in the MSB set, except that here the CLF is constrained to mass-to-light ratios for clusters of $(M/L)_{\rm cl} = 1000\ h$ (M/L)$_\odot$, rather than $(M/L)_{\rm cl} = 500\ h$ (M/L)$_\odot$ as in MSB (see Section 6.3). The MGRSs denoted by '$b_{\rm vel}$' are similar except for a velocity bias of $b_{\rm vel} = \sigma_{\rm gal}/\sigma_{\rm DM} = 0.6$ (see Section 6.2). The MGRSs denoted by 'σ_8' are also similar except that they adopt a flat ΛCDM cosmology with $\sigma_8 = 0.75$ rather than 0.9 (see Section 6.4). The final line lists the best-fitting parameters obtained by Hawkins et al. (2003) by fitting the $\xi(r_p, \pi)$ obtained from the 2dFGRS. Note that the errors in Hawkins et al. are estimated from the spread of 22 Mock samples.

[5] Note that the PVDs thus obtained are a kind of average of the true PVD along the line-of-sight. Therefore, these PVDs should not be compared directly to the true PVD shown in Fig. 9.

degeneracy on our $\sigma_{12}(r_p)$ by repeating the same exercise using a value for β that is 0.1 larger (smaller) than the values listed in Table 1. This leads to an increase (decrease) of $\sigma_{12}(r_p)$ of the order of 5 per cent (20 per cent) at projected radii of 1 h^{-1} Mpc (10 h^{-1} Mpc). Given that our MGRSs overpredict the PVD at $r_p = 1\ h^{-1}$ Mpc by about 70 per cent, it is clear that this discrepancy is not a reflection of the β–σ_{12} degeneracy. Thus, the standard ΛCDM model seems to have a severe problem in matching the observed PVDs on intermediate scales.

6 TOWARDS A SELF-CONSISTENT MODEL FOR LARGE-SCALE STRUCTURE

Our MGRSs, based on a flat ΛCDM concordance cosmology with $\Omega_{\rm m} = 0.3$ and $\sigma_8 = 0.9$, and on a CLF that is required to yield cluster mass-to-light ratios of $(M/L)_{\rm cl} = 500\,h\ ({\rm M/L})_\odot$, reveal clustering statistics that are overall in reasonable agreement with the data from the 2dFGRS. Nevertheless, two discrepancies have come to light: the MGRSs predict too much power on small scales and PVDs that are too high. We now investigate possible ways to alleviate these discrepancies.

6.1 Halo occupation models

The discrepancies between our MGRS and the 2dFGRS results might indicate a problem with our halo occupation models. Although the CLF is fairly well constrained by the observed luminosity function and the observed luminosity dependence of the correlation lengths (see Papers I and II), we have made additional assumptions regarding the second moments of the halo occupation number distributions and regarding the distribution of galaxies within individual dark matter haloes.

As we have shown in Section 4.2 the real space correlation function depends only very weakly on our method of distributing satellite galaxies within dark matter haloes (cf. Fig. 8). We have verified, using a number of tests, that modifications of the spatial distribution of satellite galaxies within dark matter haloes have no significant influence on $w_p(r_p)$ or on $\sigma_{12}(r_p)$. Therefore, none of the discrepancies mentioned above can be attributed to errors in our satellite model.

Our results are more susceptible to changes in the second moment of the halo occupation number distributions. On small scales, $\xi(r)$ scales with the average number of galaxy pairs in individual haloes $\langle \frac{1}{2}N(N-1)\rangle$. Therefore, one can decrease the power on small scales, to bring our $w_p(r_p)$ into better agreement with observations, by lowering the second moments of our halo occupation distributions. However, our distribution (16) is already the narrowest distribution possible, and modifying the second moment of the halo occupation distributions can therefore only aggravate the discrepancies on small scales.

6.2 Velocity bias

A seemingly obvious explanation for the too-high PVDs is that the peculiar velocities of galaxies are biased with respect to the dark matter. We define the velocity bias (sometimes called 'dynamical' bias) as $b_{\rm vel} = \sigma_{\rm gal}/\sigma_{\rm DM}$, with $\sigma_{\rm gal}$ and $\sigma_{\rm DM}$ the peculiar velocity dispersions of (satellite) galaxies and dark matter particles in a given halo, respectively. Note that in our fiducial MGRSs we adopt $b_{\rm vel} = 1$ (i.e. no velocity bias). Fig. 16 shows $w_p(r_p)$, $\xi(s)$, $q(s)$ and $\sigma_{12}(r_p)$ for MGRSs (with the MSB configuration of simulation boxes) in which $b_{\rm vel} = 0.6$; i.e. the velocity dispersion of satellite galaxies is only 60 per cent of that of the dark matter particles in the same halo (dashed lines). With such a pronounced velocity bias, both $q(s)$ and the PVDs, as well as β and the global value of σ_{12} listed in Table 1, are all consistent with the 2dFGRS results.

In two recent papers, Berlind et al. (2003) and Yoshikawa et al. (2003) measured the velocity bias of 'galaxies' in a smoothed particle hydrodynamics (SPH) simulation and found that $b_{\rm vel}$ decreases from $b_{\rm vel} \sim 0.9$–1.0 for haloes with $M \sim 3 \times 10^{14}\ h^{-1}\ {\rm M}_\odot$ to $b_{\rm vel} \sim 0.6$–0.8 for haloes with $M \sim 3 \times 10^{12}\ h^{-1}\ {\rm M}_\odot$. Thus, although simulations predict that low-mass haloes might have values for the velocity bias as low as $b_{\rm vel} \simeq 0.6$, required here to bring our PVDs in agreement with observations, the PVD is dominated by galaxies in massive haloes for which these same simulations apparently predict close to $b_{\rm vel} = 1$ (i.e. no velocity bias). Furthermore, the introduction of velocity bias cannot solve the excess power on small scales. After all, the real-space correlation function is independent of $b_{\rm vel}$ such that the discrepancies regarding $w_p(r_p)$ on small scales remain (see upper-left panel of Fig. 16). To make matters worse, reducing the peculiar velocities of galaxies inside dark matter haloes, increases the small-scale power in redshift space. This means that the redshift-space correlation function $\xi(s)$ actually becomes more discrepant with the 2dFGRS data (see upper-right panel of Fig. 16). Therefore, although a certain amount of velocity bias might be expected, we do not consider it a viable solution for the problems mentioned above.

6.3 Cluster mass-to-light ratios

Because the PVD is a pair-weighted statistic it is extremely sensitive to the few richest systems in the sample (i.e. Mo, Jing & Börner 1993, 1997; Zurek et al. 1994; Marzke et al. 1995; Somerville, Primack & Nolthenius 1997). The fact that the PVDs in our MGRSs are too large compared with observations therefore might indicate that either there are too many clusters of galaxies in our mock surveys (see Section 6.4 below), or that these clusters contain too many galaxies.

Our CLF was constructed under the constraint that the average mass-to-light ratio of haloes with $M \geqslant 10^{14}\ h^{-1}\ {\rm M}_\odot$ is equal to $(M/L)_{\rm cl} = 500\,h\ ({\rm M/L})_\odot$ (in the photometric $b_{\rm J}$ band). This value is motivated by the average mass-to-light ratio of clusters obtained by Fukugita, Hogan & Peebles (1998). To reduce the number of galaxies per cluster we now set $(M/L)_{\rm cl} = 1000\,h\ ({\rm M/L})_\odot$ and repeat the entire exercise: we first use the method described in Section 2 to compute the parameters of the new conditional luminosity function. This CLF is used to construct new MGRSs (using the same configuration of simulation boxes as in MSB), from which we determine the same statistics as before.

The results are listed in Table 1 and shown as dotted lines in Fig. 16. Clearly, increasing the mass-to-light ratio of clusters lowers $q(s)$ and $\sigma_{12}(r_p)$, bringing them into better agreement with the 2dFGRS results. Although the PVDs are still somewhat too high, especially at around $\sim 0.4\ h^{-1}$ Mpc, the extent of this discrepancy is similar to its 1σ variance of the eight MGRSs, indicating that this remaining difference is consistent with 'cosmic variance'. As can be seen from Table 1, both β and γ are now in much better agreement with the 2dFGRS. In addition, the reduction of the number of galaxies in clusters significantly reduces $w_p(r_p)$ at small projected separations (see Fig. 16), bringing it into good agreement with the observations.[6] A similar reduction of small-scale power is also

[6] Note that the correlation amplitude predicted with $(M/L)_{\rm cl} = 1000$

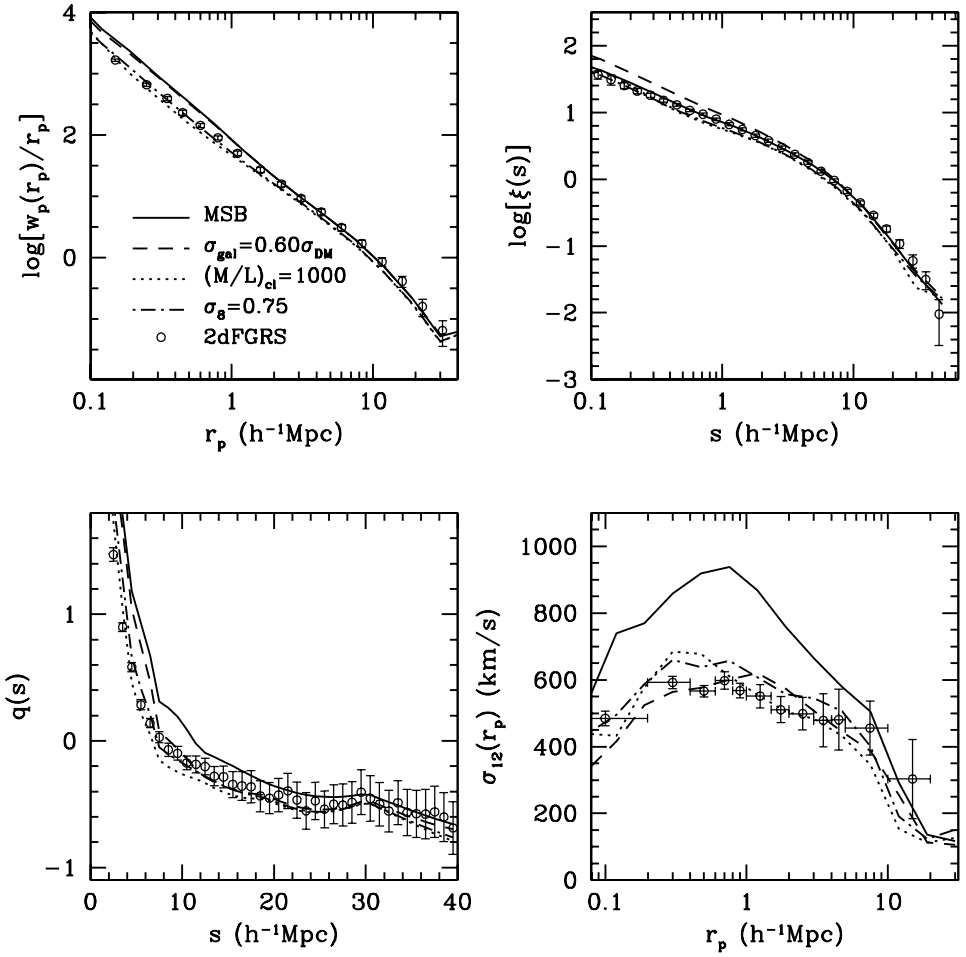

Figure 16. Same as Fig. 15 except that here we compare the results for the MSB sample with those of three alternative MGRSs in which we have modified the CLF to yield cluster mass-to-light ratios of $(M/L)_{\rm cl} = 1000\,h$ (M/L)$_\odot$ (dotted lines), in which we adopt a velocity bias of $b_{\rm vel} = 0.6$ (dashed lines), and in which we adopt a cosmology with $\sigma_8 = 0.75$ (dot-dashed lines). All results correspond to the mean of the entire sample of eight MSB mock surveys. For clarity, no error bars are plotted here, but they are similar to those shown in Fig. 15. Note that both the $(M/L)_{\rm cl} = 1000\,h$ (M/L)$_\odot$ model and the $\sigma_8 = 0.75$ model are in good agreement with the observational data.

evident in $\xi(s)$. Thus, these particular MGRSs have clustering characteristics that are overall in good agreement with the 2dFGRS results. The question is therefore whether or not such a high mass-to-light ratio for clusters of galaxies is compatible with observations.

The cluster mass-to-light ratios quoted by Fukugita et al. (1998) are (450 ± 100) h (M/L)$_\odot$ in the B-band based on X-ray and velocity-dispersion data. Taking these numbers at face value, a cluster mass-to-light ratio of $(M/L)_{\rm cl} = 1000\,h$ (M/L)$_\odot$ is ruled out at the 5σ level. Using a variety of methods to estimate cluster masses, Bahcall et al. (2000) obtained $(M/L)_B = (330 \pm 77)\,h$ (M/L)$_\odot$, which is consistent with the results of Carlberg et al. (1996), $(M/L)_B = (363 \pm 65)\,h$ (M/L)$_\odot$, based on galaxy kinematics in clusters. Taking the average of these two measurements yields $\langle M/L_B\rangle_{\rm cl} = (350 \pm 70)\,h$ (M/L)$_\odot$, which rules out the cluster mass-to-light ratio required to match the clustering power on small scales at more than 7σ. Thus, unless the cluster mass-to-light ratios obtained from current observations are seriously in error, increasing the average cluster mass-to-light ratio to $(M/L)_{\rm cl} \simeq 1000\,h$ (M/L)$_\odot$ does not seem a viable solution for the problems at hand.

6.4 Power-spectrum normalization

Rather than lowering the average number of galaxies per cluster, we may also hope to lower the PVDs by reducing the actual number of clusters. As we have shown in Section 5.2, our results, and thus the number density of rich clusters, is robust against cosmic variance. Therefore, a lower number density of clusters implies a different cosmological model. It is well known that the abundance of (rich) clusters is extremely sensitive to the power-spectrum normalization parameter σ_8. The too-high PVDs could thus be indicative of a too-high value for σ_8.

We therefore wish to compute the PVDs in a ΛCDM cosmology with identical cosmological parameters as before, except that $\sigma_8 = 0.75$ rather than 0.9. Note that the choice of $\sigma_8 = 0.75$ is somewhat arbitrary, but it does represent a compromise between the constraints on the value of σ_8 from various observations and the low value required by our results on the PVDs (see below). In principle, constructing new MGRSs for a different cosmology requires new

h(M/L)$_\odot$ is slightly lower than the observed amplitude, because in this model more galaxies are assigned to small haloes in order to match the observed luminosity function. Since the error bars on the observed correlation lengths are relatively large, the model tends to compromise the accuracy of the fit to the correlation lengths.

N-body simulations of the dark matter distribution. This, however, is computationally too expensive, which is why we use an approximate method instead. First, we compute the new best-fitting parameters of the CLF for this cosmology, again demanding that $(M/L)_{\rm cl} = 500\ h\ ({\rm M/L})_{\odot}$. Next we populate the dark matter haloes in our $\sigma_8 = 0.9$ simulation boxes with galaxies according to this new CLF. Finally, we construct a new sample of eight MSB MGRSs, in which we weigh each galaxy in a halo of mass M by

$$w = \frac{n(M\,|\,\sigma_8 = 0.75)}{n(M\,|\,\sigma_8 = 0.9)} \tag{43}$$

with $n(M)$ the number density of dark matter haloes of mass M. This, to first order, mimics the effect of lowering σ_8 on the halo mass function, and so should be a reasonable approximation on small scales where the clustering properties are determined by the galaxy distribution in individual haloes.[7] The results for these MGRSs are shown as dot-dashed lines in Fig. 16. As one can see, the agreement with observational results in this model is much better than in the standard ΛCDM model. These results suggest that a ΛCDM model with $\sigma_8 \sim 0.75$ may match all the observational results obtained from the 2dFGRS. Unfortunately, in the absence of proper N-body simulations for this model, it is impossible to make a more detailed comparison with observation.

The question is, of course, whether such a low σ_8 is compatible with other independent observations. Currently, the value of σ_8 is constrained mainly by three types of observations: weak lensing surveys, cluster mass functions, and anisotropy in the cosmic microwave background. Recent cluster abundance analyses give values of σ_8 (assuming $\Omega_{\rm m} = 0.3$) in a wide range, from 0.6 to 1 (e.g. Fan & Bahcall 1998; Pen 1998; Borgani et al. 2001; Reiprich & Böhringer 2002; Seljak 2002; Viana, Nichol & Liddle 2002). Results from weak-lensing surveys are equally uncertain, with σ_8 spanning the range 0.7 to ∼1 (e.g. Hoekstra, Yee & Gladders 2002; Refregier, Rhodes & Groth 2002; Bacon et al. 2003; Jarvis et al. 2003). Thus, our preferred value, $\sigma_8 = 0.75$, is consistent with these observations. At the moment, the most stringent constraint on the value of σ_8 is from *WMAP* (Spergel et al. 2003): $\sigma_8 = 0.84 \pm 0.04$ (1σ error). Even taking this result at face value, one cannot rule out $\sigma_8 = 0.75$ with any high confidence. Thus, there is no strong observational evidence to argue against a ΛCDM model with $\sigma_8 = 0.75$. Furthermore, as discussed by van den Bosch, Mo & Yang (2003b), a value of σ_8 as low as 0.75 can also help to alleviate several problems in current models of galaxy formation, such as those in connection with the Tully–Fisher relation and to the rotation curve shapes of low-surface-brightness galaxies. Our results presented here give additional support for a relatively low value of σ_8.

7 CONCLUSIONS

In this paper, we have used realistic halo occupation distributions, obtained using the conditional luminosity function technique introduced by Yang et al. (2003), to populate dark matter haloes in high-resolution simulations of the ΛCDM 'concordance' cosmology. The simulations follow the evolution of 512^3 dark matter particles in periodic boxes of $100\ h^{-1}$ Mpc and $300\ h^{-1}$ Mpc on a side. Subsequently, the dark matter haloes identified in these simulations are populated with galaxies of different luminosity and different morphological type.

We have shown that the luminosity functions and the correlation lengths as a function of luminosity, both for early- and late-type galaxies, are in good agreement with observations. Since these same observations were used to constrain the conditional luminosity functions, which in turn were used to populate the dark matter haloes, this agreement shows that the halo occupation statistics obtained analytically can be implemented reliably in N-body simulations to construct realistic, self-consistent, mock galaxy distributions. We have demonstrated that the details of the spatial distribution of galaxies within individual dark matter haloes have only a very mild effect on the two-point correlation function, and only at real-space separations $r \lesssim 0.3\ h^{-1}$ Mpc.

The mean pairwise peculiar velocities, $\langle v_{12}\rangle$, however, depend rather strongly on whether satellite galaxies (any galaxy in a dark matter halo other than the most luminous, central galaxy) are associated with random dark matter particles of the friends-of-friends (FOF) group, or whether they are assigned peculiar velocities assuming a spherical, isotropic velocity distribution around the central galaxy. In the former case, $\langle v_{12}\rangle$, which indicates the amount of infall around overdensity regions, is similar to that of the dark matter. In the latter case, $\langle v_{12}\rangle$ is significantly suppressed with respect to the dark matter. This difference indicates that the outer parts of the FOF-groups are not yet virialized.

The pairwise velocity dispersions (PVDs) of the galaxies are found to be significantly smaller than those of dark matter particles. Since the PVD is a pair-weighted measure for the potential well in which dark matter particles (galaxies) reside, this can be understood as long as the average number of galaxies per unit halo mass, N/M, decreases with M (Jing et al. 1998). Indeed, the halo occupation numbers inferred from our conditional luminosity function indicate that $N/M \propto M^a$ with $a \sim -0.2$.

Stacking a number of $100\ h^{-1}$ Mpc and $300\ h^{-1}$ Mpc simulation boxes allows us to construct mock galaxy redshift surveys (MGRSs) that are comparable to the 2dFGRS in terms of sky coverage, depth and magnitude limit. For each of these MGRSs we estimate the two-point correlation functions $\xi(r_p, \pi)$. These are used to derive a number of statistics about the large-scale distribution of galaxies, which we compare directly with the 2dFGRS results. In particular, we calculate the projected 2PCFs $w_p(r_p)$ as a function of luminosity and type. The best-fitting power-law slope and correlation lengths of these projected correlation functions are found in good agreement with the 2dFGRS results obtained by Norberg et al. (2002a). In addition, we also compute $w_p(r_p)$ and the redshift space correlation function $\xi(s)$ for the entire MGRSs. These are compared to the 2dFGRS results obtained by Hawkins et al. (2003).

Although the agreement with the 2dFGRS data is excellent on scales larger than $\sim 3\ h^{-1}$ Mpc, on smaller scales $w_p(r_p)$ is about a factor of 2 larger than observed. To investigate this in more detail, we analysed the redshift-space distortions present in $\xi(r_p, \pi)$ by computing the quadrupole-to-monopole ratios $q(s)$ and the pairwise velocity dispersions $\sigma_{12}(r_p)$. A comparison with the results of Hawkins et al. (2003) shows that the standard ΛCDM model overpredicts the clustering power on small scales by a factor of about 2, and the PVDs by about 350 km s^{-1}. After examining a variety of possibilities, we find that the only viable solution to these problems is to reduce the power-spectrum amplitude, σ_8, from 0.9 to ∼0.75.

No doubt, in the coming years, new results from the 2dFGRS and the SDSS will significantly improve the data on the large-scale distribution of galaxies. The analysis presented here, based on the conditional luminosity function, will hopefully prove a useful tool

[7] Note that for haloes with a given mass, the concentration parameters are smaller in the $\sigma_8 = 0.75$ model than they are in the standard ΛCDM model. This change of concentration is taken into account in our analyses, even though its effect is almost negligible.

to further constrain both galaxy formation and cosmology. In this respect, our results regarding constraints on σ_8 are an important illustration of the potential power of this approach.

ACKNOWLEDGMENTS

We thank Ed Hawkins and the 2dFGRS team for providing us with their data in electronic format, the anonymous referee for insightful comments that helped to improve the paper, and Gerhard Börner, Guinevere Kauffmann, Simon White and Saleem Zaroubi for useful discussions. XHY is supported by NSFC (No.10243005) and by USTCQN. YPJ is supported in part by NKBRSF (G19990754) and by NSFC (No.10125314). Numerical simulations presented in this paper were carried out at the Astronomical Data Analysis Centre (ADAC) of the National Astronomical Observatory, Japan.

REFERENCES

Bacon D., Massey R., Refregier A., Ellis R., 2003, MNRAS, 344, 673
Bahcall N. A., Cen R., Davé R., Ostriker J. P., Yu Q., 2000, ApJ, 541, 1
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2002, MNRAS, 316, 107
Berlind A. A., Weinberg D. H., 2002, ApJ, 575 587
Berlind A. A. et al., 2003, ApJ, 593, 1
Borgani S. et al., 2001, ApJ, 561, 13
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Bullock J. S., Wechsler, R. H., Somerville R. S., 2002, MNRAS, 329, 246
Bryan G., Norman M., 1998, ApJ, 495, 80
Carlberg R. G., Yee H. K. C., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C. J., 1996, ApJ, 462, 32
Cole S., Fisher K. B., Weinberg D. H., 1994, MNRAS, 267, 785
Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945
Colless M. et al., 2001, MNRAS, 328, 1039
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Diaferio A., Geller M. J., 1996, ApJ, 467, 19
Davis M., Huchra J., 1982, ApJ, 254, 437
Davis M., Peebles P. J. E., 1977, ApJS, 34, 425
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Fan X., Bahcall N. A., 1998, ApJ, 504, 1
Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J. P., 1994, MNRAS, 267, 927
Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518
Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, ApJ, 489, 37
Hamilton A. J. S., 1992, ApJ, 385, L5
Hamilton A. J. S., 1993, ApJ, 417, 19
Hawkins E. et al., 2003, MNRAS, 346, 78 (H03)
Hoekstra H., Yee H. K. C., Gladders M. D., 2002, ApJ, 577, 595
Huffenberger K. M., Seljak U., 2003, MNRAS, 340, 1199
Jarvis M., Bernstein G. M., Fischer P., Smith D., Jain B., Tyson J. A., Wittman D., 2003, AJ, 125, 1014
Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Jing Y. P., Börner G., Suto Y., 2002, ApJ, 564, 15
Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1
Kaiser N., 1987, MNRAS, 227, 1
Kang X., Jing Y. P., Mo H. J., Börner G., 2002, MNRAS, 336, 892
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188
Kochanek C. S., White M., Huchra J., Macri L., Jarrett T. H., Schneider S. E., Mader J., 2003, ApJ, 585, 161
Madgwick D. S. et al., 2002, MNRAS, 333, 133
Magira H., Jing Y. P., Suto Y., 2000, ApJ, 528, 30
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101
Marzke R. O., Geller M. J., Da Costa L. N., Huchra J. P., 1995, AJ, 110, 477
Mathis H., Lemson G., Springel V., Kauffmann G., White S. D. M., Eldar A., Dekel A., 2002, MNRAS, 333, 739
Mo H. J., Jing Y. P., Börner G., 1993, MNRAS, 264, 825
Mo H. J., Jing Y. P., Börner G., 1997, MNRAS, 286, 979
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Norberg P. et al., 2002a, MNRAS, 332, 827
Norberg P. et al., 2002b, MNRAS, 336, 907
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peacock J. A. et al., 2001, Nat, 410, 169
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton
Pen U.-L., 1998, ApJ, 498, 60
Porciani C., Dekel A., Hoffman Y., 2002, MNRAS, 332, 325
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Refregier A., Rhodes J., Groth E. J., 2002, ApJ, 572, L131
Regos E., Geller M. J., 1991, ApJ, 377, 14
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Scranton R., 2002, MNRAS, 332, 697
Seljak U., 2000, MNRAS, 318, 203
Seljak U., 2002, MNRAS, 337, 769
Seto N., Yokoyama J. I., 1998, ApJ, 492, 421
Sheth R. K., 1996, MNRAS, 279, 1310
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Smith R. E. et al., 2003, MNRAS, 341, 1311
Somerville R. S., Primack J. R., Nolthenius R., 1997, ApJ, 479, 606
Somerville R. S., Lemson G., Sigad Y., Dekel A., Kauffmann G., White S. D. M., 2001, MNRAS, 320, 289
Spergel D. N. et al., 2003, ApJS, 148, 175
van den Bosch F. C., Yang X. H., Mo H. J., 2003a, MNRAS, 340, 771 (Paper II)
van den Bosch F. C., Mo H. J., Yang X. H., 2003b, MNRAS, 345, 923
van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481
Viana P. P., Nichol R., Liddle A. R., 2002, ApJ, 569, L75
White M., 2001, MNRAS, 321, 1
Yang X. H., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057 (Paper I)
York D. et al., 2000, AJ, 120, 1579
Yoshikawa K., Jing Y. P., Börner G., 2003, ApJ, 590, 654
Zehavi I. et al., 2002, ApJ, 571, 172
Zehavi I. et al., 2003, preprint (astro-ph/0301280)
Zheng Z., Tinker J. L., Weinberg D. H., Berlind A. A., 2002, ApJ, 575, 617
Zurek W., Quinn P. J., Salmon T. K., Warren M. S., 1994, ApJ, 431, 559

This paper has been typeset from a TEX/LATEX file prepared by the author.